

# **Collaborative Statistics Homework Book:**

## **Custom Version modified by R. Bloom**

**Collection Editor:**

Roberta Bloom

**Authors:**

Roberta Bloom Susan Dean Barbara Illowsky, Ph.D.

**Online:**

< <http://cnx.org/content/col10619/1.2/> >

### **CONNEXIONS Rice University, Houston, Texas**

This selection and arrangement of content as a collection is copyrighted by Roberta Bloom. It is licensed under the Creative Commons Attribution 2.0 license (<http://creativecommons.org/licenses/by/2.0/>). Collection structure revised: December 23, 2009 PDF generated: May 12, 2010 For copyright and attribution information for the modules contained in this collection, see p. 228.

# Table of Contents

<b>Collaborative Statistics: custom version modified by R. Bloom .....</b>	1
<b>Preface by S. Dean and B. Illowsky .....</b>	5
<b>Additional Resources .....</b>	9
<b>Author Acknowledgements .....</b>	13
<b>Student Welcome Letter .....</b>	15
<b>1 Sampling and Data</b>	
<b>1.1 Homework (modified R. Bloom) .....</b>	17
Solutions .....	26
<b>2 Descriptive Statistics</b>	
<b>2.1 Summary of Formulas .....</b>	29
<b>2.2 Homework (modified R. Bloom) .....</b>	30
Solutions .....	46
<b>3 Probability Topics</b>	
<b>3.1 Summary of Formulas .....</b>	49
<b>3.2 Homework (modified R. Bloom) .....</b>	49
<b>3.3 Review Questions .....</b>	60
Solutions .....	63
<b>4 Discrete Random Variables</b>	
<b>4.1 Summary of Functions .....</b>	67
<b>4.2 Homework (modified R. Bloom) .....</b>	68
<b>4.3 Review Questions .....</b>	78
Solutions .....	82
<b>5 Continuous Random Variables</b>	
<b>5.1 Summary of the Uniform and Exponential Probability Distributions .....</b>	87
<b>5.2 Homework .....</b>	88
<b>5.3 Review Questions .....</b>	93
Solutions .....	96
<b>6 The Normal Distribution</b>	
<b>6.1 Summary of Formulas .....</b>	99
<b>6.2 Homework .....</b>	99
<b>6.3 Review Questions .....</b>	105
Solutions .....	107
<b>7 The Central Limit Theorem</b>	
<b>7.1 Summary of Formulas .....</b>	109
<b>7.2 Homework: CLT (modified R. Bloom) .....</b>	109
<b>7.3 Review Questions .....</b>	115
Solutions .....	118
<b>8 Confidence Intervals</b>	
<b>8.1 Summary of Formulas .....</b>	121
<b>8.2 Homework .....</b>	121
<b>8.3 Review Questions .....</b>	131
Solutions .....	135
<b>9 Hypothesis Testing: Single Mean and Single Proportion</b>	

<b>9.1</b>	Summary of Formulas .....	139
<b>9.2</b>	Homework (modified R. Bloom) .....	140
<b>9.3</b>	Review Questions .....	146
	Solutions .....	149
<b>10</b>	<b>Hypothesis Testing: Two Means, Paired Data, Two Proportions</b>	
<b>10.1</b>	Summary of Types of Hypothesis Tests .....	153
<b>10.2</b>	Homework .....	153
<b>10.3</b>	Review Questions .....	164
	Solutions .....	166
<b>11</b>	<b>The Chi-Square Distribution</b>	
<b>11.1</b>	Summary of Formulas .....	169
<b>11.2</b>	Homework .....	170
<b>11.3</b>	Review Questions .....	177
	Solutions .....	182
<b>12</b>	<b>Linear Regression and Correlation</b>	
<b>12.1</b>	95% Critical Values of the Sample Correlation Coefficient Table .....	187
<b>12.2</b>	Summary .....	188
<b>12.3</b>	Homework .....	189
	Solutions .....	204
<b>13</b>	<b>F Distribution and ANOVA</b>	
<b>13.1</b>	Summary .....	207
<b>13.2</b>	Homework .....	207
<b>13.3</b>	Review Questions .....	209
	Solutions .....	213
<b>14</b>	<b>Appendix</b>	
<b>14.1</b>	Data Sets .....	215
<b>14.2</b>	English Phrases Written Mathematically .....	219
<b>14.3</b>	Symbols and their Meanings .....	220
<b>Index</b>	.....	226
<b>Attributions</b>	.....	228

# Collaborative Statistics: custom version modified by R. Bloom<sup>1</sup>

This module summarizes the modifications made by Roberta Bloom to the modules included in the custom textbook collection Collaborative Statistics by R. Bloom <http://cnx.org/content/col10617/><sup>2</sup> <http://cnx.org/content/col10617/> and Homework Book for Collaborative Statistics by R. Bloom <http://cnx.org/content/col10619/><sup>3</sup> These custom collections are based on the textbook collection Collaborative Statistics, by Illowsky, B. and S. Dean, Connexions Web site, <http://cnx.org/content/col10522/1.29/>, Dec 5, 2008, but have been modified, as detailed in this module. If future modifications are made to the custom collection, this module will be updated to contain current information.

## **IMPORTANT NOTE TO STUDENTS OWNING A PRINT COPY OF THIS TEXTBOOK:**

This custom version of the Collaborative Statistics textbook by Susan Dean and Barbara Illowsky has been modified by Roberta Bloom. The sections that are different in this book are listed below, by title, along with a description of the changes. Section numbers and page numbers may be different also, but the section titles should correspond to the sections in the original Dean/Illowsky textbook.

**If you are using a print copy of this textbook and your class instructor is NOT Ms. Bloom:**

- you need to be aware of the textbook changes listed below
- you may need to go to Collaborative Statistics collection by S. Dean and B. Illowsky (<http://cnx.org/content/col10522/latest/>)<sup>4</sup> to view or print out the versions of the listed sections as included in the Dean/Illowsky collection for this textbook so that you have the same version as the rest of your class.

**The custom collection for Ms. Bloom's class has divided the textbook into two collections:**

- Homework Collection Collaborative Statistics Homework Collection (modified R. Bloom) (<http://cnx.org/content/col10619/latest/>)<sup>5</sup> which contains the formula summary page, the homework problems and the review problems
- Textbook Collection Collaborative Statistics (custom collection modified R. Bloom) (<http://cnx.org/content/col10617/latest/>)<sup>6</sup> which contains the text and the chapter practices, but not the homework or review problems.
- YOU NEED TO USE BOTH THE HOMEWORK COLLECTION AND THE TEXTBOOK COLLECTION FOR MS. BLOOM'S CLASS.

**List of Modifications:**

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m18941/1.3/>>.

<sup>2</sup>*Collaborative Statistics: Custom Version modified by R. Bloom* <<http://cnx.org/content/col10617/latest/>>

<sup>3</sup>*Collaborative Statistics Homework Book: Custom Version modified by R. Bloom* <<http://cnx.org/content/col10619/latest/>>

<sup>4</sup>*Collaborative Statistics* <<http://cnx.org/content/col10522/latest/>>

<sup>5</sup>*Collaborative Statistics Homework Book: Custom Version modified by R. Bloom* <<http://cnx.org/content/col10619/latest/>>

<sup>6</sup>*Collaborative Statistics: Custom Version modified by R. Bloom* <<http://cnx.org/content/col10617/latest/>>

### **Labs and Projects Removed**

The labs and projects have been removed from this modified version of Collaborative Statistics. Ms. Bloom posts the labs (and projects, if any) for her class on her class website. If you are using this book with another instructor you may need to access and print out the labs or projects online from the original Dean/Illowsky Collaborative Statistics textbook collection: <http://cnx.org/content/col10522/latest>

### **Chapter 1 Data and Sampling**

- Homework: two new homework problems have been added

### **Chapter 2 Descriptive Statistics**

- Measuring the Spread of the Data: some revisions in wording and use of symbols; introduced terminology for z-score; formulas added; brief summary of Chebyshev's Rule and Empirical Rule have been added.
- Practice 3: Interpreting Percentiles: new section added that was not included in the original textbook
- Homework: Some new homework questions have been added

### **Chapter 3: Probability Topics**

- Terminology: wording revisions; discussion of the Law of Large Numbers added
- Independent and Mutually Exclusive Events: wording revisions; add additional worked example has been added to illustrate determining that two events are not independent
- Contingency Tables: one example was removed from this section
- Practice 1: the data has been presented in tabular form
- Homework: new problems #33 through #41 have been added

### **Chapter 4: Discrete Distributions**

- Homework: new problems #38 through #43 have been added

### **Chapter 5: Continuous Probability Distributions**

- Introduction to Continuous Random Variables: some material has been removed
- Properties of Continuous Random Variables: the concepts of probability as area, including graphs illustrating this concept; probability density functions, and cumulative distribution functions are explained
- Uniform Distribution: Example 4 is new, illustrating the uniform distribution when the minimum value is not 0. This replaces the example for conditional probability that had been in the original module.
- Practice 1: Uniform Distribution: Problems pertaining to conditional probability have been removed from this section

### **Chapter 6: The Normal Distribution**

- No changes have been made to this chapter

### **Chapter 7: Central Limit Theorem**

- The section for CLT for Sums has been retitled as OPTIONAL
- Using the Central Limit Theorem : Examples illustrating use of the CLT for sums have been removed from this section.
- Practice: problems pertaining to the CLT for sums were removed from the practice

- Homework: Several homework exercises pertaining to the CLT for sums have been removed. Several homework exercises have been changed to reflect only the CLT for means: parts pertaining to the CLT for sums were replaced with new parts or removed. Specific changes made: Exercise #1 parts g,i and exercise #15 parts d,e,i and exercise #6 have been removed. Exercises #7, #11, #13, #22, #23 have been modified. Exercise #24 is new.
- Review: the context of the last problem has been changed

### **Chapter 8: Confidence Interval Estimates**

- Confidence Interval for an Unknown Population Mean, Population Standard Deviation Known: Normal Distribution: worked examples have modified to show the step by step solution using the error bound formulas to find the confidence intervals; revised module may also contain some revisions in wording.
- Confidence Interval for an Unknown Population Mean, Population Standard Deviation Known: Normal Distribution: An additional example has been added to the module illustrating how to find the mean and the error bound when only the confidence interval is given.
- Confidence Interval for an Unknown Population Mean, Population Standard Deviation Unknown: Student t Distribution: worked examples have modified to show the step by step solution using the error bound formulas to find the confidence intervals; revised module may also contain some revisions in wording.
- Confidence Interval for an Unknown Population Proportion: worked examples have modified to show the step by step solution using the error bound formulas to find the confidence intervals; revised module may also contain some revisions in wording.
- In all 3 sections listed above, the emphasis has been changed to using the error bound formulas to calculate the confidence interval rather than reliance on the calculators' interval functions to find the confidence interval.

### **Chapter 9: Hypothesis Test for a Single Mean or Single Proportion**

- Homework: Some homework exercises have been omitted in this revision; the numbering remains unchanged for the remaining exercises. The omitted exercises are indicated in the section.

### **Chapter 10: Hypothesis Testing: Two Means, Paired Data, Two Proportions**

- At this time, no changes have been made in this chapter.

### **Chapter 11: THe Chi-Square Distribution**

- No changes have been made in this chapter.

### **Chapter 12: Linear Regression and Correlation**

- Sections 12.5 (The Regression Equation), 12.6 (Correlation Coefficient and Coefficient of Determination), 12.7 (Facts about the Correlation Coefficient), 12.8 (Prediction), 12.9 Outliers have been modified
- Section 12.5 now contains calculator instructions for the LinRegTTest for the TI-83,83+,84+ calculators
- Section 12.6 includes the coefficient of determination that was not included in the original section
- Section 12.7 contains additional information about the assumptions underlying the test of significance of the correlation coefficient
- Section 12.9 includes a graphical method of identifying outliers, in addition to the numerical method included in the original version of this section
- In the Homework Section, 2 new problems have been added.

### **Chpater 13: F Distribution and ANOVA**

- No changes have been made in this chapter.

**Hypothesis Test Solution Sheets**

- Hypothesis test solution sheets for chapters 9, 10, 11, 13 will be modified and will be made available, appropriately formatted, via links on Ms. Bloom's class website.
- The original solution sheets will be removed from the online collection.

**Practice Final Exams**

- Practice Final Exams have been removed from this collection
- Instead, Practice Final Exams will be made available via links on Ms. Bloom's class website

# Preface by S. Dean and B. Illowsky<sup>7</sup>

Welcome to *Collaborative Statistics*, presented by Connexions. The initial section below introduces you to Connexions. If you are familiar with Connexions, please skip to About "Collaborative Statistics." (Section : About Connexions)

## About Connexions

### Connexions Modular Content

Connexions ([cnx.org<sup>8</sup>](http://cnx.org)) is an online, **open access** educational resource dedicated to providing high quality learning materials free online, free in printable PDF format, and at low cost in bound volumes through print-on-demand publishing. The *Collaborative Statistics* textbook is one of many **collections** available to Connexions users. Each **collection** is composed of a number of re-usable learning **modules** written in the Connexions XML markup language. Each module may also be re-used (or 're-purposed') as part of other collections and may be used outside of Connexions. Including *Collaborative Statistics*, Connexions currently offers over 6500 modules and more than 350 collections.

The modules of *Collaborative Statistics* are derived from the original paper version of the textbook under the same title, *Collaborative Statistics*. Each module represents a self-contained concept from the original work. Together, the modules comprise the original textbook.

### Re-use and Customization

The Creative Commons (CC) Attribution license<sup>9</sup> applies to all Connexions modules. Under this license, any module in Connexions may be used or modified for any purpose as long as proper attribution to the original author(s) is maintained. Connexions' authoring tools make re-use (or re-purposing) easy. Therefore, instructors anywhere are permitted to create customized versions of the *Collaborative Statistics* textbook by editing modules, deleting unneeded modules, and adding their own supplementary modules. Connexions' authoring tools keep track of these changes and maintain the CC license's required attribution to the original authors. This process creates a new collection that can be viewed online, downloaded as a single PDF file, or ordered in any quantity by instructors and students as a low-cost printed textbook. To start building custom collections, please visit the help page, "Create a Collection with Existing Modules"<sup>10</sup>. For a guide to authoring modules, please look at the help page, "Create a Module in Minutes"<sup>11</sup>.

### Read the book online, print the PDF, or buy a copy of the book.

To browse the *Collaborative Statistics* textbook online, visit the collection home page at [cnx.org/content/col10522/latest<sup>12</sup>](http://cnx.org/content/col10522/latest). You will then have three options.

<sup>7</sup>This content is available online at <<http://cnx.org/content/m16026/1.16/>>.

<sup>8</sup><http://cnx.org/>

<sup>9</sup><http://creativecommons.org/licenses/by/2.0/>

<sup>10</sup><http://cnx.org/help/CreateCollection>

<sup>11</sup><http://cnx.org/help/ModuleInMinutes>

<sup>12</sup>*Collaborative Statistics* <<http://cnx.org/content/col10522/latest>>

1. You may obtain a PDF of the entire textbook to print or view offline by clicking on the "Download PDF" link in the "Content Actions" box.
2. You may order a bound copy of the collection by clicking on the "Order Printed Copy" button.
3. You may view the collection modules online by clicking on the "Start >>" link, which takes you to the first module in the collection. You can then navigate through the subsequent modules by using their "Next >>" and "Previous >>" links to move forward and backward in the collection. You can jump to any module in the collection by clicking on that module's title in the "Collection Contents" box on the left side of the window. If these contents are hidden, make them visible by clicking on "[show table of contents]".

### **Accessibility and Section 508 Compliance**

- For information on general Connexions accessibility features, please visit <http://cnx.org/content/m17212/latest/><sup>13</sup>.
- For information on accessibility features specific to the Collaborative Statistics textbook, please visit <http://cnx.org/content/m17211/latest/><sup>14</sup>.

### **Version Change History and Errata**

- For a list of modifications, updates, and corrections, please visit <http://cnx.org/content/m17360/latest/><sup>15</sup>.

### **Adoption and Usage**

- The Collaborative Statistics collection has been adopted and customized by a number of professors and educators for use in their classes. For a list of known versions and adopters, please visit <http://cnx.org/content/m18261/latest/><sup>16</sup>.

## **About "Collaborative Statistics"**

*Collaborative Statistics* was written by Barbara Illowsky and Susan Dean, faculty members at De Anza College in Cupertino, California. The textbook was developed over several years and has been used in regular and honors-level classroom settings and in distance learning classes. Courses using this textbook have been articulated by the University of California for transfer of credit. The textbook contains full materials for course offerings, including expository text, examples, labs, homework, and projects. A Teacher's Guide is currently available in print form and on the Connexions site at <http://cnx.org/content/col10547/latest/><sup>17</sup>, and supplemental course materials including additional problem sets and video lectures are available at <http://cnx.org/content/col10586/latest/><sup>18</sup>. The on-line text for each of these collections will meet the Section 508 standards for accessibility.

An on-line course based on the textbook was also developed by Illowsky and Dean. It has won an award as the best on-line California community college course. The on-line course will be available at a later date as a collection in Connexions, and each lesson in the on-line course will be linked to the on-line textbook chapter. The on-line course will include, in addition to expository text and examples, videos of course lectures in captioned and non-captioned format.

The original preface to the book as written by professors Illowsky and Dean, now follows:

---

<sup>13</sup>"Accessibility Features of Connexions" <<http://cnx.org/content/m17212/latest/>>

<sup>14</sup>"Collaborative Statistics: Accessibility" <<http://cnx.org/content/m17211/latest/>>

<sup>15</sup>"Collaborative Statistics: Change History" <<http://cnx.org/content/m17360/latest/>>

<sup>16</sup>"Collaborative Statistics: Adoption and Usage" <<http://cnx.org/content/m18261/latest/>>

<sup>17</sup>*Collaborative Statistics Teacher's Guide* <<http://cnx.org/content/col10547/latest/>>

<sup>18</sup>*Collaborative Statistics: Supplemental Course Materials* <<http://cnx.org/content/col10586/latest/>>

This book is intended for introductory statistics courses being taken by students at two- and four-year colleges who are majoring in fields other than math or engineering. Intermediate algebra is the only prerequisite. The book focuses on applications of statistical knowledge rather than the theory behind it. The text is named *Collaborative Statistics* because students learn best by **doing**. In fact, they learn best by working in small groups. The old saying “two heads are better than one” truly applies here.

**Our emphasis in this text is on four main concepts:**

- thinking statistically
- incorporating technology
- working collaboratively
- writing thoughtfully

These concepts are integral to our course. Students learn the best by actively participating, not by just watching and listening. Teaching should be highly interactive. Students need to be thoroughly engaged in the learning process in order to make sense of statistical concepts. *Collaborative Statistics* provides techniques for students to write across the curriculum, to collaborate with their peers, to think statistically, and to incorporate technology.

This book takes students step by step. The text is interactive. Therefore, students can immediately apply what they read. Once students have completed the process of problem solving, they can tackle interesting and challenging problems relevant to today’s world. The problems require the students to apply their newly found skills. In addition, technology (TI-83 graphing calculators are highlighted) is incorporated throughout the text and the problems, as well as in the special group activities and projects. The book also contains labs that use real data and practices that lead students step by step through the problem solving process.

At De Anza, along with hundreds of other colleges across the country, the college audience involves a large number of ESL students as well as students from many disciplines. The ESL students, as well as the non-ESL students, have been especially appreciative of this text. They find it extremely readable and understandable. *Collaborative Statistics* has been used in classes that range from 20 to 120 students, and in regular, honor, and distance learning classes.

Susan Dean

Barbara Illowsky

# Additional Resources<sup>19</sup>

## Additional Resources Currently Available

- Glossary (Glossary, p. 9)
- View or Download This Textbook Online (View or Download This Textbook Online, p. 9)
- Collaborative Statistics Teacher's Guide (Collaborative Statistics Teacher's Guide, p. 9)
- Supplemental Materials (Supplemental Materials, p. 9)
- Video Lectures (Video Lectures, p. 10)
- Version History (Version History, p. 10)
- Textbook Adoption and Usage (Textbook Adoption and Usage, p. 10)
- Additional Technologies and Notes (Additional Technologies, p. 10)
- Accessibility and Section 508 Compliance (Accessibility and Section 508 Compliance, p. 10)

The following section describes some additional resources for learners and educators. These modules and collections are all available on the Connexions website (<http://cnx.org/><sup>20</sup>) and can be viewed online, downloaded, printed, or ordered as appropriate.

### Glossary

This module contains the entire glossary for the Collaborative Statistics textbook collection (col10522) since its initial release on 15 July 2008. The glossary is located at <http://cnx.org/content/m16129/latest/><sup>21</sup>.

### View or Download This Textbook Online

The complete contents of this book are available at no cost on the Connexions website at <http://cnx.org/content/col10522/latest/><sup>22</sup>. Anybody can view this content free of charge either as an online e-book or a downloadable PDF file. A low-cost printed version of this textbook is also available here<sup>23</sup>.

### Collaborative Statistics Teacher's Guide

A complementary Teacher's Guide for Collaborative statistics is available through Connexions at <http://cnx.org/content/col10547/latest/><sup>24</sup>. The Teacher's Guide includes suggestions for presenting concepts found throughout the book as well as recommended homework assignments. A low-cost printed version of this textbook is also available here<sup>25</sup>.

### Supplemental Materials

This companion to Collaborative Statistics provides a number of additional resources for use by students and instructors based on the award winning Elementary Statistics Sofia online course<sup>26</sup>, also by textbook

<sup>19</sup>This content is available online at <<http://cnx.org/content/m18746/1.5/>>.

<sup>20</sup><http://cnx.org/>

<sup>21</sup>"Collaborative Statistics: Glossary" <<http://cnx.org/content/m16129/latest/>>

<sup>22</sup>*Collaborative Statistics* <<http://cnx.org/content/col10522/latest/>>

<sup>23</sup><http://my.qoop.com/store/7064943342106149/7781159220340>

<sup>24</sup>*Collaborative Statistics Teacher's Guide* <<http://cnx.org/content/col10547/latest/>>

<sup>25</sup><http://my.qoop.com/store/7064943342106149/8791310589747>

<sup>26</sup><http://sofia.flda.edu/gallery/statistics/index.html>

authors Barbara Illowsky and Susan Dean. This content is designed to complement the textbook by providing video tutorials, course management materials, and sample problem sets. The Supplemental Materials collection can be found at <http://cnx.org/content/col10586/latest/><sup>27</sup>.

### Video Lectures

- Video Lecture 1: Sampling and Data<sup>28</sup>
- Video Lecture 2: Descriptive Statistics<sup>29</sup>
- Video Lecture 3: Probability Topics<sup>30</sup>
- Video Lecture 4: Discrete Distributions<sup>31</sup>
- Video Lecture 5: Continuous Random Variables<sup>32</sup>
- Video Lecture 6: The Normal Distribution<sup>33</sup>
- Video Lecture 7: The Central Limit Theorem<sup>34</sup>
- Video Lecture 8: Confidence Intervals<sup>35</sup>
- Video Lecture 9: Hypothesis Testing with a Single Mean<sup>36</sup>
- Video Lecture 10: Hypothesis Testing with Two Means<sup>37</sup>
- Video Lecture 11: The Chi-Square Distribution<sup>38</sup>
- Video Lecture 12: Linear Regression and Correlation<sup>39</sup>

### Version History

This module contains a listing of changes, updates, and corrections made to the Collaborative Statistics textbook collection (col10522) since its initial release on 15 July 2008. The Version History is located at <http://cnx.org/content/m17360/latest/><sup>40</sup>.

### Textbook Adoption and Usage

This module is designed to track the various derivations of the Collaborative Statistics textbook and its various companion resources, as well as keep track of educators who have adopted various versions for their courses. New adopters are encouraged to provide their contact information and describe how they will use this book for their courses. The goal is to provide a list that will allow educators using this book to collaborate, share ideas, and make suggestions for future development of this text. The Adoption and Usage module is located at <http://cnx.org/content/m18261/latest/><sup>41</sup>.

### Additional Technologies

In order to provide the most flexible learning resources possible, we invite collaboration from all instructors wishing to create customized versions of this content for use with other technologies. For instance, you may be interested in creating a set of instructions similar to this collection's calculator notes. If you would like to contribute to this collection, please use the contact the authors with any ideas or materials you have created.

### Accessibility and Section 508 Compliance

---

<sup>27</sup>Collaborative Statistics: Supplemental Course Materials <<http://cnx.org/content/col10586/latest/>>

<sup>28</sup>"Elementary Statistics: Video Lecture - Sampling and Data" <<http://cnx.org/content/m17561/latest/>>

<sup>29</sup>"Elementary Statistics: Video Lecture - Descriptive Statistics" <<http://cnx.org/content/m17562/latest/>>

<sup>30</sup>"Elementary Statistics: Video Lecture - Probability Topics" <<http://cnx.org/content/m17563/latest/>>

<sup>31</sup>"Elementary Statistics: Video Lecture - Discrete Distributions" <<http://cnx.org/content/m17565/latest/>>

<sup>32</sup>"Elementary Statistics: Video Lecture - Continuous Random Variables" <<http://cnx.org/content/m17566/latest/>>

<sup>33</sup>"Elementary Statistics: Video Lecture - The Normal Distribution" <<http://cnx.org/content/m17567/latest/>>

<sup>34</sup>"Elementary Statistics: Video Lecture - The Central Limit Theorem" <<http://cnx.org/content/m17568/latest/>>

<sup>35</sup>"Elementary Statistics: Video Lecture - Confidence Intervals" <<http://cnx.org/content/m17569/latest/>>

<sup>36</sup>"Elementary Statistics: Video Lecture - Hypothesis Testing with a Single Mean" <<http://cnx.org/content/m17570/latest/>>

<sup>37</sup>"Elementary Statistics: Video Lecture - Hypothesis Testing with Two Means" <<http://cnx.org/content/m17577/latest/>>

<sup>38</sup>"Elementary Statistics: Video Lecture - The Chi-Square Distribution" <<http://cnx.org/content/m17571/latest/>>

<sup>39</sup>"Elementary Statistics: Video Lecture - Linear Regression and Correlation" <<http://cnx.org/content/m17572/latest/>>

<sup>40</sup>"Collaborative Statistics: Change History" <<http://cnx.org/content/m17360/latest/>>

<sup>41</sup>"Collaborative Statistics: Adoption and Usage" <<http://cnx.org/content/m18261/latest/>>

- For information on general Connexions accessibility features, please visit <http://cnx.org/content/m17212/latest/><sup>42</sup>.
- For information on accessibility features specific to the Collaborative Statistics textbook, please visit <http://cnx.org/content/m17211/latest/><sup>43</sup>.

---

<sup>42</sup>"Accessibility Features of Connexions" <<http://cnx.org/content/m17212/latest/>>

<sup>43</sup>"Collaborative Statistics: Accessibility" <<http://cnx.org/content/m17211/latest/>>

# Author Acknowledgements <sup>44</sup>

We wish to acknowledge the many people who have helped us and have encouraged us in this project. At De Anza, Donald Rossi and Rupinder Sekhon and their contagious enthusiasm started us on our path to this book. Inna Grushko and Diane Mathios painstakingly checked every practice and homework problem. Inna also wrote the glossary and offered invaluable suggestions.

Kathy Plum co-taught with us the first term we introduced the TI-85. Lenore Desilets, Charles Klein, Kathy Plum, Janice Hector, Vernon Paige, Carol Olmstead, and Donald Rossi of De Anza College, Ann Flanigan of Kapiolani Community College, Birgit Aquilonius of West Valley College, and Terri Teegarden of San Diego Mesa College, graciously volunteered to teach out of our early editions. Janice Hector and Lenore Desilets also contributed problems. Diane Mathios and Carol Olmstead contributed labs as well. In addition, Diane and Kathy have been our “sounding boards” for new ideas. In recent years, Lisa Markus, Vladimir Logvinenko, and Roberta Bloom have contributed valuable suggestions.

Jim Lucas and Valerie Hauber of De Anza’s Office of Institutional Research, along with Mary Jo Kane of Health Services, provided us with a wealth of data.

We would also like to thank the thousands of students who have used this text. So many of them gave us permission to include their outstanding word problems as homework. They encouraged us to turn our note packet into this book, have offered suggestions and criticisms, and keep us going.

Finally, we owe much to Frank, Jeffrey, and Jessica Dean and to Dan, Rachel, Matthew, and Rebecca Ilowsky, who encouraged us to continue with our work and who had to hear more than their share of “I’m sorry, I can’t” and “Just a minute, I’m working.”

---

<sup>44</sup>This content is available online at <<http://cnx.org/content/m16308/1.9/>>.

# Student Welcome Letter<sup>45</sup>

Dear Student:

Have you heard others say, "You're taking statistics? That's the hardest course I ever took!" They say that, because they probably spent the entire course confused and struggling. They were probably lectured to and never had the chance to experience the subject. You will not have that problem. Let's find out why.

There is a Chinese Proverb that describes our feelings about the field of statistics:

I HEAR, AND I FORGET

I SEE, AND I REMEMBER

I DO, AND I UNDERSTAND

Statistics is a "do" field. In order to learn it, you must "do" it. We have structured this book so that you will have hands-on experiences. They will enable you to truly understand the concepts instead of merely going through the requirements for the course.

What makes this book different from other texts? First, we have eliminated the drudgery of tedious calculations. You might be using computers or graphing calculators so that you do not need to struggle with algebraic manipulations. Second, this course is taught as a collaborative activity. With others in your class, you will work toward the common goal of learning this material.

Here are some hints for success in your class:

- Work hard and work every night.
- Form a study group and learn together.
- Don't get discouraged - you can do it!
- As you solve problems, ask yourself, "Does this answer make sense?"
- Many statistics words have the same meaning as in everyday English.
- Go to your teacher for help as soon as you need it.
- Don't get behind.
- Read the newspaper and ask yourself, "Does this article make sense?"
- Draw pictures - they truly help!

Good luck and don't give up!

Sincerely,  
Susan Dean and Barbara Illowsky

---

<sup>45</sup>This content is available online at <<http://cnx.org/content/m16305/1.5/>>.



# Chapter 1

## Sampling and Data

### 1.1 Homework (modified R. Bloom)<sup>1</sup>

#### Exercise 1.1

For each item below:

(*Solution on p. 26.*)

- i. Identify the type of data (quantitative - discrete, quantitative - continuous, or qualitative) that would be used to describe a response.
- ii. Give an example of the data.
  - a. Number of tickets sold to a concert
  - b. Amount of body fat
  - c. Favorite baseball team
  - d. Time in line to buy groceries
  - e. Number of students enrolled at Evergreen Valley College
  - f. Most-watched television show
  - g. Brand of toothpaste
  - h. Distance to the closest movie theatre
  - i. Age of executives in Fortune 500 companies
  - j. Number of competing computer spreadsheet software packages

#### Exercise 1.2

Fifty part-time students were asked how many courses they were taking this term. The (incomplete) results are shown below:

Part-time Student Course Loads

# of Courses	Frequency	Relative Frequency	Cumulative Frequency	Relative Frequency
continued on next page				

<sup>1</sup>This content is available online at <<http://cnx.org/content/m18858/1.3/>>.

1	30	0.6	
2	15		
3			

**Table 1.1**

- a. Fill in the blanks in the table above.
- b. What percent of students take exactly two courses?
- c. What percent of students take one or two courses?

**Exercise 1.3***(Solution on p. 26.)*

Sixty adults with gum disease were asked the number of times per week they used to floss before their diagnoses. The (incomplete) results are shown below:

**Flossing Frequency for Adults with Gum Disease**

# Flossing per Week	Frequency	Relative Frequency	Cumulative Relative Freq.
0	27	0.4500	
1	18		
3			0.9333
6	3	0.0500	
7	1	0.0167	

**Table 1.2**

- a. Fill in the blanks in the table above.
- b. What percent of adults flossed six times per week?
- c. What percent flossed at most three times per week?

**Exercise 1.4**

A fitness center is interested in the average amount of time a client exercises in the center each week. Define the following in terms of the study. Give examples where appropriate.

- a. Population
- b. Sample
- c. Parameter
- d. Statistic
- e. Variable
- f. Data

**Exercise 1.5***(Solution on p. 26.)*

Ski resorts are interested in the average age that children take their first ski and snowboard lessons. They need this information to optimally plan their ski classes. Define the following in terms of the study. Give examples where appropriate.

- a. Population
- b. Sample
- c. Parameter

- d. Statistic
- e. Variable
- f. Data

**Exercise 1.6**

A cardiologist is interested in the average recovery period for her patients who have had heart attacks. Define the following in terms of the study. Give examples where appropriate.

- a. Population
- b. Sample
- c. Parameter
- d. Statistic
- e. Variable
- f. Data

**Exercise 1.7**

*(Solution on p. 26.)*

Insurance companies are interested in the average health costs each year for their clients, so that they can determine the costs of health insurance. Define the following in terms of the study. Give examples where appropriate.

- a. Population
- b. Sample
- c. Parameter
- d. Statistic
- e. Variable
- f. Data

**Exercise 1.8**

A politician is interested in the proportion of voters in his district that think he is doing a good job. Define the following in terms of the study. Give examples where appropriate.

- a. Population
- b. Sample
- c. Parameter
- d. Statistic
- e. Variable
- f. Data

**Exercise 1.9**

*(Solution on p. 26.)*

A marriage counselor is interested in the proportion the clients she counsels that stay married. Define the following in terms of the study. Give examples where appropriate.

- a. Population
- b. Sample
- c. Parameter
- d. Statistic
- e. Variable
- f. Data

**Exercise 1.10**

Political pollsters may be interested in the proportion of people that will vote for a particular cause. Define the following in terms of the study. Give examples where appropriate.

- a. Population
- b. Sample
- c. Parameter
- d. Statistic
- e. Variable
- f. Data

**Exercise 1.11***(Solution on p. 26.)*

A marketing company is interested in the proportion of people that will buy a particular product. Define the following in terms of the study. Give examples where appropriate.

- a. Population
- b. Sample
- c. Parameter
- d. Statistic
- e. Variable
- f. Data

**Exercise 1.12**

Airline companies are interested in the consistency of the number of babies on each flight, so that they have adequate safety equipment. Suppose an airline conducts a survey. Over Thanksgiving weekend, it surveys 6 flights from Boston to Salt Lake City to determine the number of babies on the flights. It determines the amount of safety equipment needed by the result of that study.

- a. Using complete sentences, list three things wrong with the way the survey was conducted.
- b. Using complete sentences, list three ways that you would improve the survey if it were to be repeated.

**Exercise 1.13**

Suppose you want to determine the average number of students per statistics class in your state. Describe a possible sampling method in 3 – 5 complete sentences. Make the description detailed.

**Exercise 1.14**

Suppose you want to determine the average number of cans of soda drunk each month by persons in their twenties. Describe a possible sampling method in 3 - 5 complete sentences. Make the description detailed.

**Exercise 1.15**

726 distance learning students at Long Beach City College in the 2004-2005 academic year were surveyed and asked the reasons they took a distance learning class. (*Source: Amit Schitai, Director of Instructional Technology and Distance Learning, LBCC*). The results of this survey are listed in the table below.

### Reasons for Taking LBCC Distance Learning Courses

Convenience	87.6%
Unable to come to campus	85.1%
Taking on-campus courses in addition to my DL course	71.7%
Instructor has a good reputation	69.1%
To fulfill requirements for transfer	60.8%
To fulfill requirements for Associate Degree	53.6%
Thought DE would be more varied and interesting	53.2%
I like computer technology	52.1%
Had success with previous DL course	52.0%
On-campus sections were full	42.1%
To fulfill requirements for vocational certification	27.1%
Because of disability	20.5%

**Table 1.3**

Assume that the survey allowed students to choose from the responses listed in the table above.

- a. Why can the percents add up to over 100%?
- b. Does that necessarily imply a mistake in the report?
- c. How do you think the question was worded to get responses that totaled over 100%?
- d. How might the question be worded to get responses that totaled 100%?

#### **Exercise 1.16**

Nineteen immigrants to the U.S were asked how many years, to the nearest year, they have lived in the U.S. The data are as follows:

2; 5; 7; 2; 2; 10; 20; 15; 0; 7; 0; 20; 5; 12; 15; 12; 4; 5; 10

The following table was produced:

#### **Frequency of Immigrant Survey Responses**

Data	Frequency	Relative Frequency	Cumulative Relative Frequency
0	2	$\frac{2}{19}$	0.1053
2	3	$\frac{3}{19}$	0.2632
4	1	$\frac{1}{19}$	0.3158
5	3	$\frac{3}{19}$	0.1579
7	2	$\frac{2}{19}$	0.5789
10	2	$\frac{2}{19}$	0.6842
12	2	$\frac{2}{19}$	0.7895
15	1	$\frac{1}{19}$	0.8421
20	1	$\frac{1}{19}$	1.0000

**Table 1.4**

- a. Fix the errors on the table. Also, explain how someone might have arrived at the incorrect number(s).
- b. Explain what is wrong with this statement: “47 percent of the people surveyed have lived in the U.S. for 5 years.”
- c. Fix the statement above to make it correct.
- d. What fraction of the people surveyed have lived in the U.S. 5 or 7 years?
- e. What fraction of the people surveyed have lived in the U.S. at most 12 years?
- f. What fraction of the people surveyed have lived in the U.S. fewer than 12 years?
- g. What fraction of the people surveyed have lived in the U.S. from 5 to 20 years, inclusive?

**Exercise 1.17**

A “random survey” was conducted of 3274 people of the “microprocessor generation” (people born since 1971, the year the microprocessor was invented). It was reported that 48% of those individuals surveyed stated that if they had \$2000 to spend, they would use it for computer equipment. Also, 66% of those surveyed considered themselves relatively savvy computer users. (*Source: San Jose Mercury News*)

- a. Do you consider the sample size large enough for a study of this type? Why or why not?
- b. Based on your “gut feeling,” do you believe the percents accurately reflect the U.S. population for those individuals born since 1971? If not, do you think the percents of the population are actually higher or lower than the sample statistics? Why?

Additional information: The survey was reported by Intel Corporation of individuals who visited the Los Angeles Convention Center to see the Smithsonian Institute’s road show called “America’s Smithsonian.”

- c. With this additional information, do you feel that all demographic and ethnic groups were equally represented at the event? Why or why not?
- d. With the additional information, comment on how accurately you think the sample statistics reflect the population parameters.

**Exercise 1.18**

- a. List some practical difficulties involved in getting accurate results from a telephone survey.
- b. List some practical difficulties involved in getting accurate results from a mailed survey.
- c. With your classmates, brainstorm some ways to overcome these problems if you needed to conduct a phone or mail survey.

**1.1.1 Try these multiple choice questions**

**The next four questions refer to the following:** A Lake Tahoe Community College instructor is interested in the average number of days Lake Tahoe Community College math students are absent from class during a quarter.

**Exercise 1.19**

What is the population she is interested in?

(*Solution on p. 27.*)

- A. All Lake Tahoe Community College students

- B. All Lake Tahoe Community College English students
- C. All Lake Tahoe Community College students in her classes
- D. All Lake Tahoe Community College math students

**Exercise 1.20**

(Solution on p. 27.)

Consider the following:

$X$  = number of days a Lake Tahoe Community College math student is absent

In this case,  $X$  is an example of a:

- A. Variable
- B. Population
- C. Statistic
- D. Data

**Exercise 1.21**

(Solution on p. 27.)

The instructor takes her sample by gathering data on 5 randomly selected students from each Lake Tahoe Community College math class. The type of sampling she used is

- A. Cluster sampling
- B. Stratified sampling
- C. Simple random sampling
- D. Convenience sampling

**Exercise 1.22**

(Solution on p. 27.)

The instructor's sample produces an average number of days absent of 3.5 days. This value is an example of a

- A. Parameter
- B. Data
- C. Statistic
- D. Variable

**The next two questions** refer to the following relative frequency table on hurricanes that have made direct hits on the U.S between 1851 and 2004. Hurricanes are given a strength category rating based on the minimum wind speed generated by the storm. ([http://www.nhc.noaa.gov/gifs/table5.gif\[?\]](http://www.nhc.noaa.gov/gifs/table5.gif[?]))

**Frequency of Hurricane Direct Hits**

Category	Number of Direct Hits	Relative Frequency	Cumulative Frequency
1	109	0.3993	0.3993
2	72	0.2637	0.6630
3	71	0.2601	
4	18		0.9890
5	3	0.0110	1.0000
	<b>Total = 273</b>		

Table 1.5

**Exercise 1.23**

(Solution on p. 27.)

What is the relative frequency of direct hits that were category 4 hurricanes?

- A. 0.0768
- B. 0.0659
- C. 0.2601
- D. Not enough information to calculate

**Exercise 1.24**

(Solution on p. 27.)

What is the relative frequency of direct hits that were AT MOST a category 3 storm?

- A. 0.3480
- B. 0.9231
- C. 0.2601
- D. 0.3370

**The next three questions refer to the following:** A study was done to determine the age, number of times per week and the duration (amount of time) of resident use of a local park in San Jose. The first house in the neighborhood around the park was selected randomly and then every 8th house in the neighborhood around the park was interviewed.

**Exercise 1.25**

(Solution on p. 27.)

'Number of times per week' is what type of data?

- A. qualitative
- B. quantitative - discrete
- C. quantitative - continuous

**Exercise 1.26**

(Solution on p. 27.)

The sampling method was:

- A. simple random
- B. systematic
- C. stratified
- D. cluster

**Exercise 1.27**

(Solution on p. 27.)

'Duration (amount of time)' is what type of data?

- A. qualitative
- B. quantitative - discrete
- C. quantitative - continuous

**Exercise 1.28**

(Solution on p. 27.)

Name the sampling method used in each of the following situations:

- a. A woman in the airport is handing out questionnaires to travelers asking them to evaluate the airport's service. She does not ask travelers who are hurrying through the airport with their hands full of luggage, but instead asks all travelers sitting near gates and who are not taking naps while they wait.
- b. A teacher wants to know if her students are doing homework so she randomly selects rows 2 and 5, and then calls on all students in row 2 and all students in row 5 to present the solution to homework problems to the class.

- c. The marketing manager for an electronics chain store wants information about the ages of its customers. Over the next two weeks, at each store location, 100 randomly selected customers are given questionnaires to fill out which asks for information about age, as well as about other variables of interest.
- d. The librarian at a public library wants to determine what proportion of the library users are children. The librarian has a tally sheet on which she marks whether the books are checked out by an adult or a child. She records this data for every 4th patron who checks out books.
- e. A political party wants to know the reaction of voters to a debate between the candidates. The day after the debate, the party's polling staff calls 1200 randomly selected phone numbers. If a registered voter answers the phone or is available to come to the phone, that registered voter is asked who he/she intends to vote for and whether the debate changed his/her opinion of the candidates.

**Exercise 1.29***(Solution on p. 27.)*

Several online textbook retailers advertise that they have lower prices than on-campus bookstores. However, an important factor is whether the internet retailers actually have the textbooks that students need in stock. Students need to be able to get textbooks promptly at the beginning of the college term. If the book is not available, then a student would not be able to get the textbook at all, or might get a delayed delivery if the book is back ordered.

A college newspaper reporter is investigating textbook availability at online retailers. He decides to investigate one textbook for each of the following 7 subjects: calculus, biology, chemistry, physics, statistics, geology, and general engineering. He consults textbook industry sales data and selects the most popular nationally used textbook in each of these subjects. He visits websites for a random sample of major online textbook sellers and looks up each of these 7 textbooks to see if they are available in stock for quick delivery through these retailers. Based on his investigation, he writes an article in which he draws conclusions about the overall availability of all college textbooks through online textbook retailers.

Write an analysis of his study that addresses the following issues: Is his sample representative of the population of all college textbooks? Explain why or why not. Describe some possible sources of bias in this study, and how it might affect the results of the study. Give some suggestions about what could be done to improve the study.

## Solutions to Exercises in Chapter 1

### Solution to Exercise 1.1 (p. 17)

- a. quantitative - discrete
- b. quantitative - continuous
- c. qualitative
- d. quantitative - continuous
- e. quantitative - discrete
- f. qualitative
- g. qualitative
- h. quantitative - continuous
- i. quantitative - continuous
- j. quantitative - discrete

### Solution to Exercise 1.3 (p. 18)

- b. 5.00%
- c. 93.33%

### Solution to Exercise 1.5 (p. 18)

- a. Children who take ski and snowboard lessons
- b. A group of these children
- c. The population average
- d. The sample average
- e.  $X =$  the age of one child who takes the first ski or snowboard lesson
- f. A value for  $X$ , such as 3, 7, etc.

### Solution to Exercise 1.7 (p. 19)

- a. The clients of the insurance companies
- b. A group of the clients
- c. The average yearly health costs of all health insurance clients
- d. The average yearly health costs of the clients in the sample
- e.  $X =$  the health insurance cost for a year for one client
- f. A value for  $X$ , such as \$140, \$725, \$2143 etc.

### Solution to Exercise 1.9 (p. 19)

- a. All the clients of the counselor
- b. A group of the clients
- c. The proportion of all her clients who stay married
- d. The proportion of the sample who stay married
- e.  $X =$  the number of couples who stay married
- f. yes, no

### Solution to Exercise 1.11 (p. 20)

- a. All people (maybe in a certain geographic area, such as the United States)
- b. A group of the people
- c. The proportion of all people who will buy the product
- d. The proportion of the sample who will buy the product
- e.  $X =$  the number of people who will buy it
- f. buy, not buy

**Solution to Exercise 1.19 (p. 22)**

D

**Solution to Exercise 1.20 (p. 23)**

A

**Solution to Exercise 1.21 (p. 23)**

B

**Solution to Exercise 1.22 (p. 23)**

C

**Solution to Exercise 1.23 (p. 24)**

B

**Solution to Exercise 1.24 (p. 24)**

B

**Solution to Exercise 1.25 (p. 24)**

B

**Solution to Exercise 1.26 (p. 24)**

B

**Solution to Exercise 1.27 (p. 24)**

C

**Solution to Exercise 1.28 (p. 24)**

- a. Convenience
- b. Cluster
- c. Stratified
- d. Systematic
- e. Simple Random

**Solution to Exercise 1.29 (p. 25)**

The answer below contains some of the issues that students might discuss for this problem. Individual student's answers may also identify other issues that pertain to this problem that are not included in the answer below.

The sample is not representative of the population of all college textbooks. Two reasons why it is not representative are that he only sampled 7 subjects and he only investigated one textbook in each subject.

There are several possible sources of bias in the study. The 7 subjects that he investigated are all in mathematics and the sciences; there are many subjects in the humanities, social sciences, and many other subject areas, (for example: literature, art, history, psychology, sociology, business) that he did not investigate at all. It may be that different subject areas exhibit different patterns of textbook availability, but his sample would not detect such results. He also only looked at the most popular textbook in each of the subjects he investigated. The availability of the most popular textbooks may differ from the availability of other textbooks in one of two ways: (a) the most popular textbooks may be more readily available online, because more new copies are printed and more students nationwide selling back their used copies OR (b) the most popular textbooks may be harder to find available online, because more student demand exhausts the supply more quickly. In reality, many college students do not use the most popular textbook in their subject, and this study gives no useful information about the situation for those less popular textbooks.

He could improve this study by (a) expanding the selection of subjects he investigates so that it is more representative of all subjects studied by college students and (b) expanding the selection of textbooks he investigates within each subject to include a mixed representation of both the popular and less popular textbooks.



## Chapter 2

# Descriptive Statistics

### 2.1 Summary of Formulas<sup>1</sup>

#### Commonly Used Symbols

- The symbol  $\Sigma$  means to add or to find the sum.
- $n$  = the number of data values in a sample
- $N$  = the number of people, things, etc. in the population
- $\bar{x}$  = the sample mean
- $s$  = the sample standard deviation
- $\mu$  = the population mean
- $\sigma$  = the population standard deviation
- $f$  = frequency
- $x$  = numerical value

#### Commonly Used Expressions

- $x * f$  = A value multiplied by its respective frequency
- $\sum x$  = The sum of the values
- $\sum x * f$  = The sum of values multiplied by their respective frequencies
- $(x - \bar{x})$  or  $(x - \mu)$  = Deviations from the mean (how far a value is from the mean)
- $(x - \bar{x})^2$  or  $(x - \mu)^2$  = Deviations squared
- $f(x - \bar{x})^2$  or  $f(x - \mu)^2$  = The deviations squared and multiplied by their frequencies

#### Mean Formulas:

- $\bar{x} = \frac{\sum x}{n}$  or  $\bar{x} = \frac{\sum f \cdot x}{\sum f}$
- $\mu = \frac{\sum x}{N}$  or  $\mu = \frac{\sum f \cdot x}{N}$

#### Standard Deviation Formulas:

- $s = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}}$  or  $s = \sqrt{\frac{\sum f \cdot (x - \bar{x})^2}{n-1}}$
- $\sigma = \sqrt{\frac{\sum(x - \mu)^2}{N}}$  or  $\sigma = \sqrt{\frac{\sum f \cdot (x - \mu)^2}{N}}$

#### Formulas Relating a Value, the Mean, and the Standard Deviation:

- value = mean + (#ofSTDEVs)(standard deviation), where #ofSTDEVs = the number of standard deviations

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m16310/1.8/>>.

- $x = \bar{x} + (\# \text{of STDEVs})(s)$
- $x = \mu + (\# \text{of STDEVs})(\sigma)$

## 2.2 Homework (modified R. Bloom)<sup>2</sup>

### Exercise 2.1

*(Solution on p. 46.)*

Twenty-five randomly selected students were asked the number of movies they watched the previous week. The results are as follows:

# of movies	Frequency	Relative Frequency	Cumulative Relative Frequency
0	5		
1	9		
2	6		
3	4		
4	1		

Table 2.1

- Find the sample mean  $\bar{x}$ .
- Find the sample standard deviation,  $s$ .
- Construct a histogram of the data.
- Complete the columns of the chart.
- Find the first quartile.
- Find the median.
- Find the third quartile.
- Construct a box plot of the data.
- What percent of the students saw fewer than three movies?
- Find the 40th percentile.
- Find the 90th percentile.

### Exercise 2.2

The median age for U.S. blacks currently is 30.1 years; for U.S. whites it is 36.6 years. (Source: U.S. Census)

- Based upon this information, give two reasons why the black median age could be lower than the white median age.
- Does the lower median age for blacks necessarily mean that blacks die younger than whites? Why or why not?
- How might it be possible for blacks and whites to die at approximately the same age, but for the median age for whites to be higher?

### Exercise 2.3

*(Solution on p. 46.)*

Forty randomly selected students were asked the number of pairs of sneakers they owned. Let  $X$  = the number of pairs of sneakers owned. The results are as follows:

<sup>2</sup>This content is available online at <<http://cnx.org/content/m18645/1.4/>>.

X	Frequency	Relative Frequency	Cumulative Relative Frequency
1	2		
2	5		
3	8		
4	12		
5	12		
7	1		

**Table 2.2**

- a. Find the sample mean  $\bar{x}$
- b. Find the sample standard deviation,  $s$
- c. Construct a histogram of the data.
- d. Complete the columns of the chart.
- e. Find the first quartile.
- f. Find the median.
- g. Find the third quartile.
- h. Construct a box plot of the data.
- i. What percent of the students owned at least five pairs?
- j. Find the 40th percentile.
- k. Find the 90th percentile.

**Exercise 2.4**

600 adult Americans were asked by telephone poll, What do you think constitutes a middle-class income? The results are below. Also, include left endpoint, but not the right endpoint. (Source: *Time magazine; survey by Yankelovich Partners, Inc.*)

NOTE: "Not sure" answers were omitted from the results.

Salary (\$)	Relative Frequency
< 20,000	0.02
20,000 - 25,000	0.09
25,000 - 30,000	0.19
30,000 - 40,000	0.26
40,000 - 50,000	0.18
50,000 - 75,000	0.17
75,000 - 99,999	0.02
100,000+	0.01

**Table 2.3**

- a. What percent of the survey answered "not sure" ?
- b. What percent think that middle-class is from \$25,000 - \$50,000 ?
- c. Construct a histogram of the data

- a. Should all bars have the same width, based on the data? Why or why not?
- b. How should the <20,000 and the 100,000+ intervals be handled? Why?
- d. Find the 40th and 80th percentiles

**Exercise 2.5**

(Solution on p. 46.)

Following are the published weights (in pounds) of all of the team members of the San Francisco 49ers from a previous year (Source: San Jose Mercury News).

177; 205; 210; 210; 232; 205; 185; 185; 178; 210; 206; 212; 184; 174; 185; 242; 188; 212; 215; 247; 241; 223; 220; 260; 245; 259; 278; 270; 280; 295; 275; 285; 290; 272; 273; 280; 285; 286; 200; 215; 185; 230; 250; 241; 190; 260; 250; 302; 265; 290; 276; 228; 265

- a. Organize the data from smallest to largest value.
- b. Find the median.
- c. Find the first quartile.
- d. Find the third quartile.
- e. Construct a box plot of the data.
- f. The middle 50% of the weights are from \_\_\_\_\_ to \_\_\_\_\_.
- g. If our population were all professional football players, would the above data be a sample of weights or the population of weights? Why?
- h. If our population were the San Francisco 49ers, would the above data be a sample of weights or the population of weights? Why?
- i. Assume the population was the San Francisco 49ers. Find:
  - i. the population mean,  $\mu$ .
  - ii. the population standard deviation,  $\sigma$ .
  - iii. the weight that is 2 standard deviations below the mean.
  - iv. When Steve Young, quarterback, played football, he weighed 205 pounds. How many standard deviations above or below the mean was he?
- j. That same year, the average weight for the Dallas Cowboys was 240.08 pounds with a standard deviation of 44.38 pounds. Emmitt Smith weighed in at 209 pounds. With respect to his team, who was lighter, Smith or Young? How did you determine your answer?
- k. Based on the shape of the data, what is the most appropriate measure of center for this data: mean, median, or mode? Explain.
- l. Are there any outliers in the data? Use an appropriate numerical test involving the IQR to identify outliers, if any, and clearly state your conclusion.
- m. Are any data values further away than 2 standard deviations from the mean? Clearly state your conclusion and show numerical work to justify your answer.

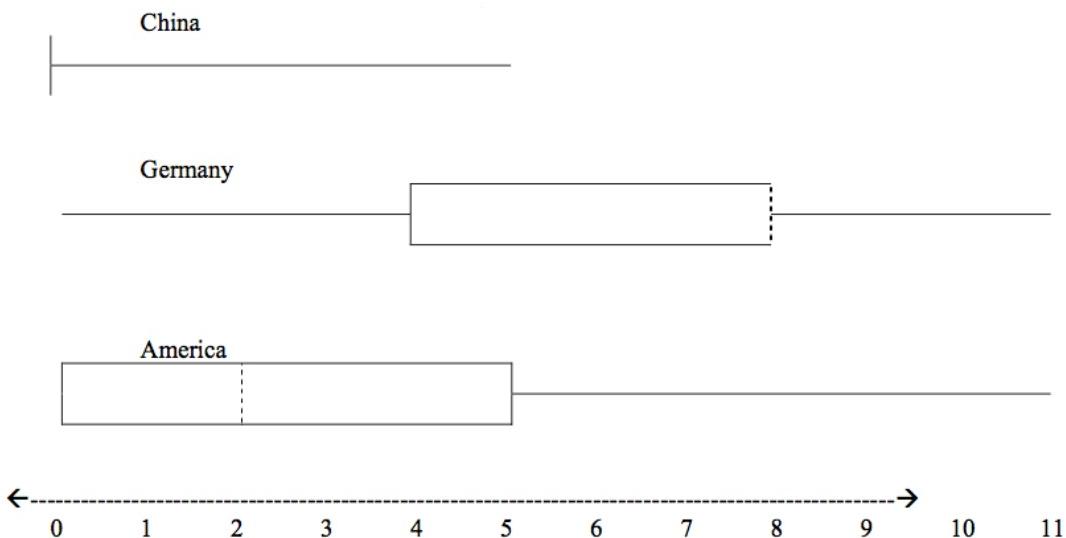
**Exercise 2.6**

An elementary school class ran 1 mile in an average of 11 minutes with a standard deviation of 3 minutes. Rachel, a student in the class, ran 1 mile in 8 minutes. A junior high school class ran 1 mile in an average of 9 minutes, with a standard deviation of 2 minutes. Kenji, a student in the class, ran 1 mile in 8.5 minutes. A high school class ran 1 mile in an average of 7 minutes with a standard deviation of 4 minutes. Nedda, a student in the class, ran 1 mile in 8 minutes.

- a. Why is Kenji considered a better runner than Nedda, even though Nedda ran faster than he?
- b. Who is the fastest runner with respect to his or her class? Explain why.

**Exercise 2.7**

In a survey of 20 year olds in China, Germany and America, people were asked the number of foreign countries they had visited in their lifetime. The following box plots display the results.



- In complete sentences, describe what the shape of each box plot implies about the distribution of the data collected.
- Explain how it is possible that more Americans than Germans surveyed have been to over eight foreign countries.
- Compare the three box plots. What do they imply about the foreign travel of twenty year old residents of the three countries when compared to each other?

### Exercise 2.8

Twelve teachers attended a seminar on mathematical problem solving. Their attitudes were measured before and after the seminar. A positive number change attitude indicates that a teacher's attitude toward math became more positive. The twelve change scores are as follows:

3; 8; -1; 2; 0; 5; -3; 1; -1; 6; 5; -2

- What is the average change score?
- What is the standard deviation for this population?
- What is the median change score?
- Find the change score that is 2.2 standard deviations below the mean.

### Exercise 2.9

*(Solution on p. 47.)*

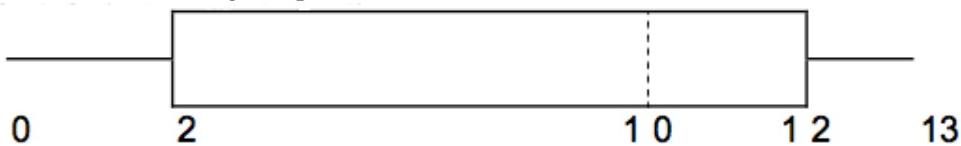
Three students were applying to the same graduate school. They came from schools with different grading systems. Which student had the best G.P.A. when compared to his school? Explain how you determined your answer.

Student	G.P.A.	School Ave. G.P.A.	School Standard Deviation
Thuy	2.7	3.2	0.8
Vichet	87	75	20
Kamala	8.6	8	0.4

Table 2.4

**Exercise 2.10**

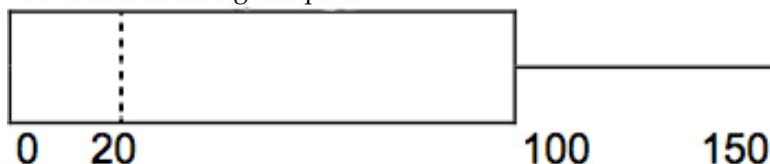
Given the following box plot:



- Which quarter has the smallest spread of data? What is that spread?
  - Which quarter has the largest spread of data? What is that spread?
  - Find the Inter Quartile Range (IQR).
  - Are there more data in the interval 5 - 10 or in the interval 10 - 13? How do you know this?
  - Which interval has the fewest data in it? How do you know this?
- I. 0-2
  - II. 2-4
  - III. 10-12
  - IV. 12-13

**Exercise 2.11**

Given the following box plot:



- Think of an example (in words) where the data might fit into the above box plot. In 2-5 sentences, write down the example.
- What does it mean to have the first and second quartiles so close together, while the second to fourth quartiles are far apart?

**Exercise 2.12**

Santa Clara County, CA, has approximately 27,873 Japanese-Americans. Their ages are as follows. (Source: West magazine)

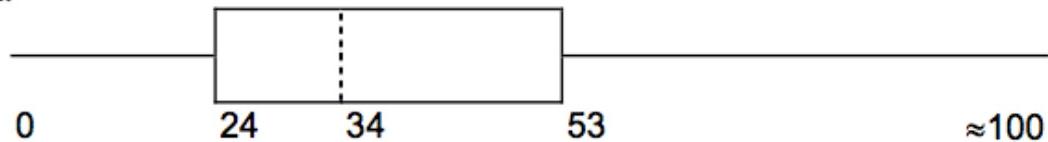
Age Group	Percent of Community
0-17	18.9
18-24	8.0
25-34	22.8
35-44	15.0
45-54	13.1
55-64	11.9
65+	10.3

Table 2.5

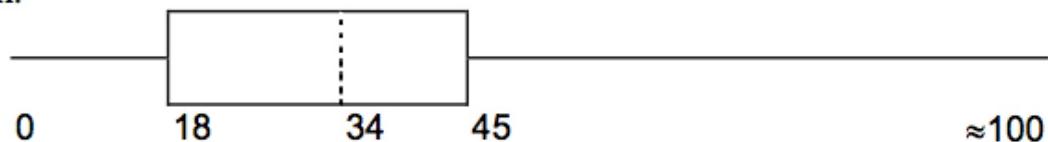
- Construct a histogram of the Japanese-American community in Santa Clara County, CA. The bars will **not** be the same width for this example. Why not?

- b. What percent of the community is under age 35?  
 c. Which box plot most resembles the information above?

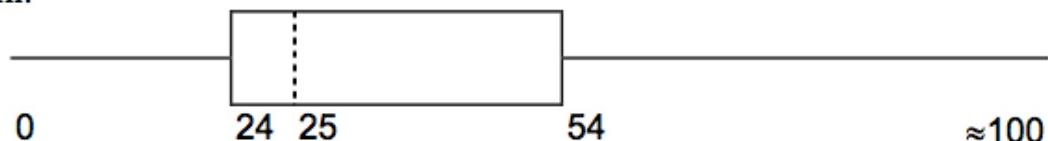
i.



ii.



iii.



### Exercise 2.13

Suppose that three book publishers were interested in the number of fiction paperbacks adult consumers purchase per month. Each publisher conducted a survey. In the survey, each asked adult consumers the number of fiction paperbacks they had purchased the previous month. The results are below.

**Publisher A**

# of books	Freq.	Rel. Freq.
0	10	
1	12	
2	16	
3	12	
4	8	
5	6	
6	2	
8	2	

**Table 2.6**

**Publisher B**

# of books	Freq.	Rel. Freq.
0	18	
1	24	
2	24	
3	22	
4	15	
5	10	
7	5	
9	1	

**Table 2.7****Publisher C**

# of books	Freq.	Rel. Freq.
0-1	20	
2-3	35	
4-5	12	
6-7	2	
8-9	1	

**Table 2.8**

- a. Find the relative frequencies for each survey. Write them in the charts.
- b. Using either a graphing calculator, computer, or by hand, use the frequency column to construct a histogram for each publisher's survey. For Publishers A and B, make bar widths of 1. For Publisher C, make bar widths of 2.
- c. In complete sentences, give two reasons why the graphs for Publishers A and B are not identical.
- d. Would you have expected the graph for Publisher C to look like the other two graphs? Why or why not?
- e. Make new histograms for Publisher A and Publisher B. This time, make bar widths of 2.
- f. Now, compare the graph for Publisher C to the new graphs for Publishers A and B. Are the graphs more similar or more different? Explain your answer.

**Exercise 2.14**

Often, cruise ships conduct all on-board transactions, with the exception of gambling, on a cash-less basis. At the end of the cruise, guests pay one bill that covers all on-board transactions. Suppose that 60 single travelers and 70 couples were surveyed as to their on-board bills for a seven-day cruise from Los Angeles to the Mexican Riviera. Below is a summary of the bills for each group.

**Singles**

Amount(\$)	Frequency	Rel. Frequency
51-100	5	
101-150	10	
151-200	15	
201-250	15	
251-300	10	
301-350	5	

**Table 2.9****Couples**

Amount(\$)	Frequency	Rel. Frequency
100-150	5	
201-250	5	
251-300	5	
301-350	5	
351-400	10	
401-450	10	
451-500	10	
501-550	10	
551-600	5	
601-650	5	

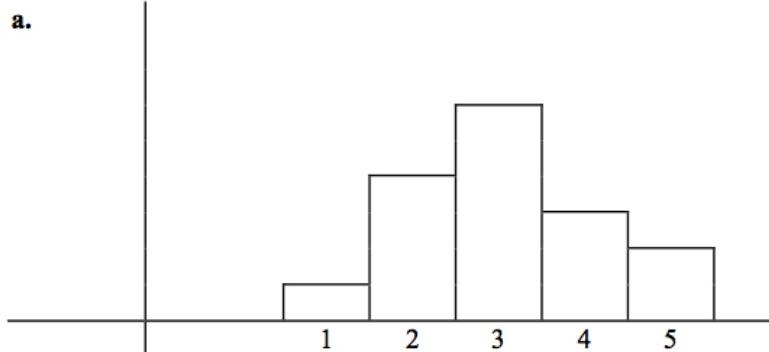
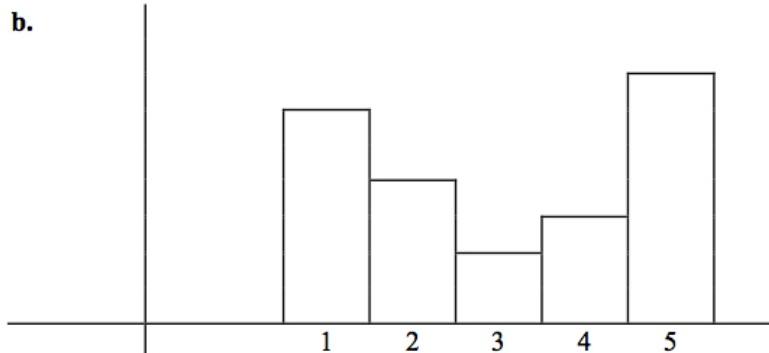
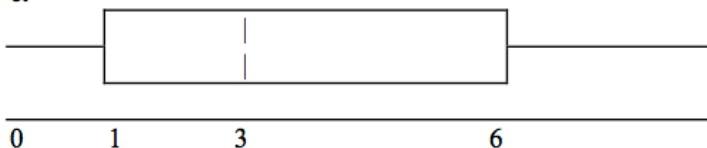
**Table 2.10**

- a. Fill in the relative frequency for each group.
- b. Construct a histogram for the Singles group. Scale the x-axis by \$50. widths. Use relative frequency on the y-axis.
- c. Construct a histogram for the Couples group. Scale the x-axis by \$50. Use relative frequency on the y-axis.
- d. Compare the two graphs:
  - i. List two similarities between the graphs.
  - ii. List two differences between the graphs.
  - iii. Overall, are the graphs more similar or different?
- e. Construct a new graph for the Couples by hand. Since each couple is paying for two individuals, instead of scaling the x-axis by \$50, scale it by \$100. Use relative frequency on the y-axis.
- f. Compare the graph for the Singles with the new graph for the Couples:
  - i. List two similarities between the graphs.
  - ii. Overall, are the graphs more similar or different?

- i. By scaling the Couples graph differently, how did it change the way you compared it to the Singles?
- j. Based on the graphs, do you think that individuals spend the same amount, more or less, as singles as they do person by person in a couple? Explain why in one or two complete sentences.

**Exercise 2.15***(Solution on p. 47.)*

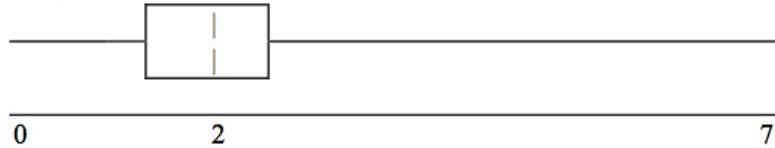
Refer to the following histograms and box plot. Determine which of the following are true and which are false. Explain your solution to each part in complete sentences.

**a.****b.****c.**

- a. The medians for all three graphs are the same.
- b. We cannot determine if any of the means for the three graphs is different.
- c. The standard deviation for (b) is larger than the standard deviation for (a).
- d. We cannot determine if any of the third quartiles for the three graphs is different.

**Exercise 2.16**

Refer to the following box plots.

**Data 1****Data 2**

- In complete sentences, explain why each statement is false.
  - Data 1** has more data values above 2 than **Data 2** has above 2.
  - The data sets cannot have the same mode.
  - For **Data 1**, there are more data values below 4 than there are above 4.
- For which group, Data 1 or Data 2, is the value of "7" more likely to be an outlier? Explain why in complete sentences

**Exercise 2.17**

(Solution on p. 47.)

In a recent issue of the *IEEE Spectrum*, 84 engineering conferences were announced. Four conferences lasted two days. Thirty-six lasted three days. Eighteen lasted four days. Nineteen lasted five days. Four lasted six days. One lasted seven days. One lasted eight days. One lasted nine days. Let  $X$  = the length (in days) of an engineering conference.

- Organize the data in a chart.
- Find the median, the first quartile, and the third quartile.
- Find the 65th percentile.
- Find the 10th percentile.
- Construct a box plot of the data.
- The middle 50% of the conferences last from \_\_\_\_\_ days to \_\_\_\_\_ days.
- Calculate the sample mean of days of engineering conferences.
- Calculate the sample standard deviation of days of engineering conferences.
- Find the mode.
- If you were planning an engineering conference, which would you choose as the length of the conference: mean; median; or mode? Explain why you made that choice.
- Give two reasons why you think that 3 - 5 days seem to be popular lengths of engineering conferences.

**Exercise 2.18**

A survey of enrollment at 35 community colleges across the United States yielded the following figures (source: Microsoft Bookshelf):

6414; 1550; 2109; 9350; 21828; 4300; 5944; 5722; 2825; 2044; 5481; 5200; 5853; 2750; 10012; 6357; 27000; 9414; 7681; 3200; 17500; 9200; 7380; 18314; 6557; 13713; 17768; 7493; 2771; 2861; 1263; 7285; 28165; 5080; 11622

- Organize the data into a chart with five intervals of equal width. Label the two columns "Enrollment" and "Frequency."
- Construct a histogram of the data.

- c. If you were to build a new community college, which piece of information would be more valuable: the mode or the average size?
- d. Calculate the sample average.
- e. Calculate the sample standard deviation.
- f. A school with an enrollment of 8000 would be how many standard deviations away from the mean?

**Exercise 2.19**

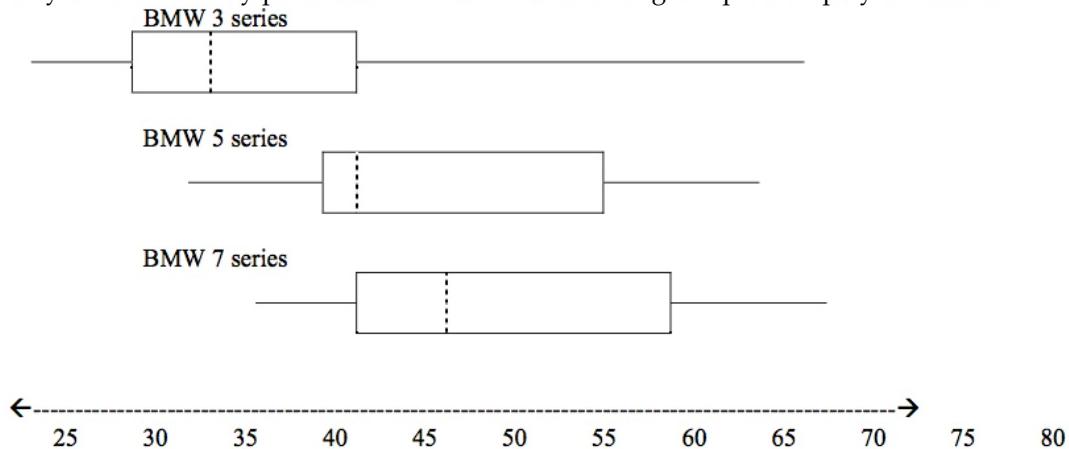
(Solution on p. 47.)

The median age of the U.S. population in 1980 was 30.0 years. In 1991, the median age was 33.1 years. (Source: Bureau of the Census)

- a. What does it mean for the median age to rise?
- b. Give two reasons why the median age could rise.
- c. For the median age to rise, is the actual number of children less in 1991 than it was in 1980? Why or why not?

**Exercise 2.20**

A survey was conducted of 130 purchasers of new BMW 3 series cars, 130 purchasers of new BMW 5 series cars, and 130 purchasers of new BMW 7 series cars. In it, people were asked the age they were when they purchased their car. The following box plots display the results.

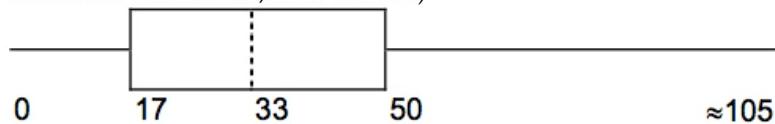


- a. In complete sentences, describe what the shape of each box plot implies about the distribution of the data collected for that car series.
- b. Which group is most likely to have an outlier? Explain how you determined that.
- c. Compare the three box plots. What do they imply about the age of purchasing a BMW from the series when compared to each other?
- d. Look at the BMW 5 series. Which quarter has the smallest spread of data? What is that spread?
- e. Look at the BMW 5 series. Which quarter has the largest spread of data? What is that spread?
- f. Look at the BMW 5 series. Find the Inter Quartile Range (IQR).
- g. Look at the BMW 5 series. Are there more data in the interval 31-38 or in the interval 45-55? How do you know this?
- h. Look at the BMW 5 series. Which interval has the fewest data in it? How do you know this?
  - i. 31-35
  - ii. 38-41
  - iii. 41-64

**Exercise 2.21**

(Solution on p. 47.)

The following box plot shows the U.S. population for 1990, the latest available year. (Source: Bureau of the Census, 1990 Census)



- Are there fewer or more children (age 17 and under) than senior citizens (age 65 and over)? How do you know?
- 12.6% are age 65 and over. Approximately what percent of the population are of working age adults (above age 17 to age 65)?

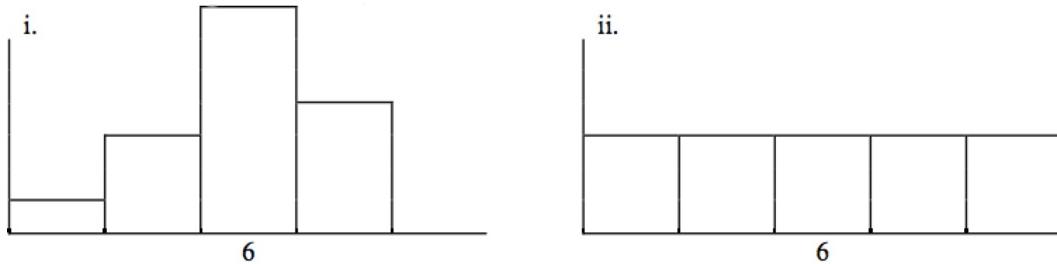
**Exercise 2.22**

Javier and Ercilia are supervisors at a shopping mall. Each was given the task of estimating the mean distance that shoppers live from the mall. They each randomly surveyed 100 shoppers. The samples yielded the following information:

	Javier	Ercilia
$\bar{x}$	6.0 miles	6.0 miles
$s$	4.0 miles	7.0 miles

**Table 2.11**

- How can you determine which survey was correct?
- Explain what the difference in the results of the surveys implies about the data.
- If the two histograms depict the distribution of values for each supervisor, which one depicts Ercilia's sample? How do you know?

**Figure 2.1**

- If the two box plots depict the distribution of values for each supervisor, which one depicts Ercilia's sample? How do you know?

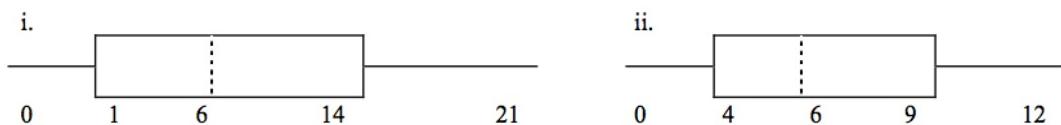


Figure 2.2

**Exercise 2.23**

(Solution on p. 47.)

Student grades on a chemistry exam were:

77, 78, 76, 81, 86, 51, 79, 82, 84, 99

- Construct a stem-and-leaf plot of the data.
- Are there any potential outliers? If so, which scores are they? Why do you consider them outliers?

**2.2.1 Try these multiple choice questions.**

The next three questions refer to the following information. We are interested in the number of years students in a particular elementary statistics class have lived in California. The information in the following table is from the entire section.

Number of years	Frequency
7	1
14	3
15	1
18	1
19	4
20	3
22	1
23	1
26	1
40	2
42	2
<b>Total = 20</b>	

Table 2.12

**Exercise 2.24**

(Solution on p. 47.)

What is the IQR?

- A. 8

- B. 11
- C. 15
- D. 35

**Exercise 2.25***(Solution on p. 47.)*

What is the mode?

- A. 19
- B. 19.5
- C. 14 and 20
- D. 22.65

**Exercise 2.26***(Solution on p. 47.)*

Is this a sample or the entire population?

- A. sample
- B. entire population
- C. neither

**The next two questions refer to the following table.** X = the number of days per week that 100 clients use a particular exercise facility.

X	Frequency
0	3
1	12
2	33
3	28
4	11
5	9
6	4

**Table 2.13****Exercise 2.27***(Solution on p. 48.)*

The 80th percentile is:

- A. 5
- B. 80
- C. 3
- D. 4

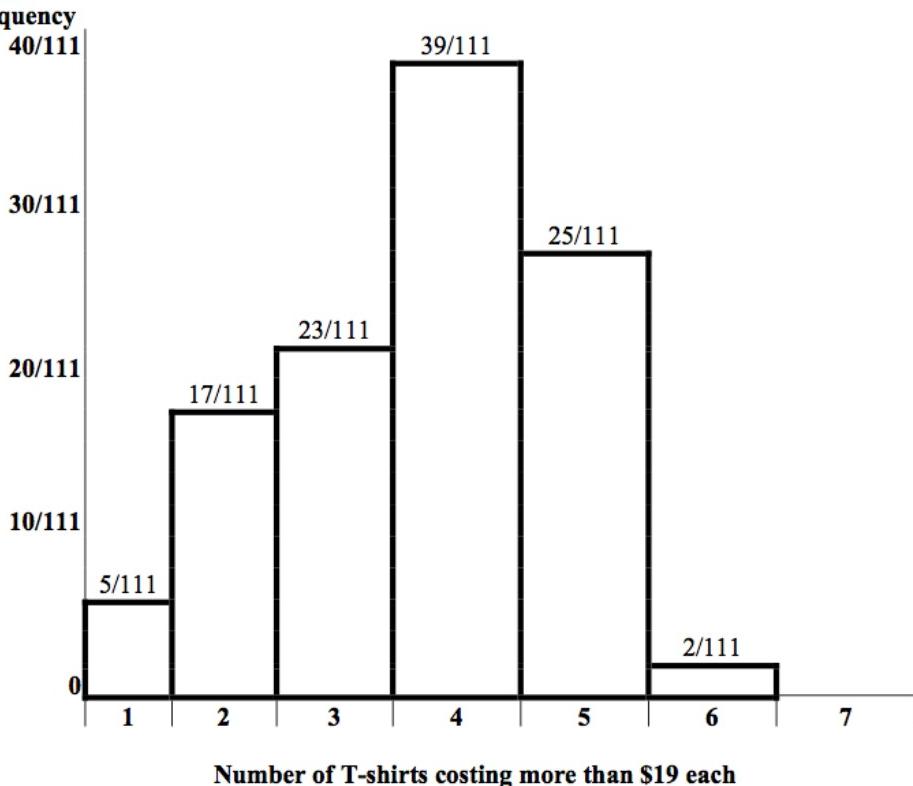
**Exercise 2.28***(Solution on p. 48.)*

The number that is 1.5 standard deviations BELOW the mean is approximately:

- A. 0.7
- B. 4.8
- C. -2.8
- D. Cannot be determined

The next two questions refer to the following histogram. Frederico recently opened a "designer" T-shirt store near the beach. During the first month of operation, he conducted a marketing survey of a random sample of 111 customers. One of the questions asked the customer how many T-shirts he/she owns that cost more than \$19 each.

**Relative Frequency**



**Exercise 2.29**

(Solution on p. 48.)

The percent of people that own at most three (3) T-shirts costing more than \$19 each is approximately:

- A. 21
- B. 59
- C. 41
- D. Cannot be determined

**Exercise 2.30**

(Solution on p. 48.)

If the data were collected by asking the first 111 people who entered the store, then the type of sampling is:

- A. cluster
- B. simple random
- C. stratified
- D. convenience

**Exercise 2.31**

(Solution on p. 48.)

A music school has budgeted to purchase 3 musical instruments. They plan to purchase a piano costing \$3000, a guitar costing \$550, and a drum set costing \$600. The average cost for a piano is \$4,000 with a standard deviation of \$2,500. The average cost for a guitar is \$500 with a standard

deviation of \$200. The average cost for drums is \$700 with a standard deviation of \$100. Which cost is the lowest, when compared to other instruments of the same type? Which cost is the highest when compared to other instruments of the same type. Justify your answer numerically.

**Exercise 2.32**

*(Solution on p. 48.)*

Suppose that a publisher conducted a survey asking adult consumers the number of fiction paperback books they had purchased in the previous month. The results are summarized in the table below. (Note that this is the data presented for publisher B in homework exercise 13).

**Publisher B**

# of books	Freq.	Rel. Freq.
0	18	
1	24	
2	24	
3	22	
4	15	
5	10	
7	5	
9	1	

**Table 2.14**

- a. Are there any outliers in the data? Use an appropriate numerical test involving the IQR to identify outliers, if any, and clearly state your conclusion.
- b. If a data value is identified as an outlier, what should be done about it?
- c. Are any data values further than 2 standard deviations away from the mean? In some situations, statisticians may use this criteria to identify data values that are unusual, compared to the other data values. (Note that this criteria is most appropriate to use for data that is mound-shaped and symmetric, rather than for skewed data.)
- d. Do parts (a) and (c) of this problem give the same answer?
- e. Examine the shape of the data. Which part, (a) or (c), of this question gives a more appropriate result for this data?
- f. Based on the shape of the data which is the most appropriate measure of center for this data: mean, median or mode?

## Solutions to Exercises in Chapter 2

### Solution to Exercise 2.1 (p. 30)

- a. 1.48
- b. 1.12
- e. 1
- f. 1
- g. 2



- h.  $\overline{0 \quad 1 \quad 2 \quad 3 \quad 4}$

- i. 80%
- j. 1
- k. 3

### Solution to Exercise 2.3 (p. 30)

- a. 3.78
- b. 1.29
- e. 3
- f. 4
- g. 5

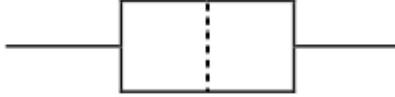


- h.  $\overline{1 \quad 3 \quad 4 \quad 5 \quad 7}$

- i. 32.5%
- j. 4
- k. 5

### Solution to Exercise 2.5 (p. 32)

- b. 241
- c. 205.5
- d. 272.5



- e.  $\overline{174 \quad 205.5 \quad 241 \quad 272.5 \quad 302}$

- f. 205.5, 272.5
- g. sample
- h. population
- i. i. 236.34  
ii. 37.50

iii. 161.34

iv. 0.84 std. dev. below the mean

j. Young

k. The mean is most appropriate. From the boxplot the data appear to be relatively symmetric. When the data are symmetric, it is appropriate to use the mean because it incorporates more information from the data. (If the data were skewed, then it would be more appropriate to use the median; but these data are not skewed.)

l.  $IQR = 272.5 - 202.5 = 70$ ;  $Q1 - 1.5 \cdot IQR = 205.5 - 1.5(70) = 105$ ;  $Q3 + 1.5 \cdot IQR = 272.5 + 1.5(70) = 373$ . All weights are between 105 and 373. There are no outliers.

m. Mean  $- 2(\text{standard deviation}) = 240.08 - 2(44.38) = 151.32$ ; Mean  $+ 2(\text{standard deviation}) = 240.08 + 2(44.38) = 328.84$ ; All players' weights are between 2 standard deviations above and below the mean.

### Solution to Exercise 2.9 (p. 33)

Kamala

### Solution to Exercise 2.15 (p. 38)

a. True

b. True

c. True

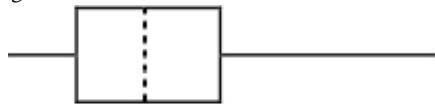
d. False

### Solution to Exercise 2.17 (p. 39)

b. 4,3,5

c. 4

d. 3



e. 2    3    4    5

f. 3,5

g. 3.94

h. 1.28

i. 3

j. mode

### Solution to Exercise 2.19 (p. 40)

c. Maybe

### Solution to Exercise 2.21 (p. 41)

a. more children

b. 62.4%

### Solution to Exercise 2.23 (p. 42)

b. 51,99

### Solution to Exercise 2.24 (p. 42)

A

### Solution to Exercise 2.25 (p. 43)

A

**Solution to Exercise 2.26 (p. 43)**

B

**Solution to Exercise 2.27 (p. 43)**

D

**Solution to Exercise 2.28 (p. 43)**

A

**Solution to Exercise 2.29 (p. 44)**

C

**Solution to Exercise 2.30 (p. 44)**

D

**Solution to Exercise 2.31 (p. 44)**

For pianos, the cost of the piano is 0.4 standard deviations BELOW average. For guitars, the cost of the guitar is 0.25 standard deviations ABOVE average. For drums, the cost of the drum set is 1.0 standard deviations BELOW average. Of the three, the drums cost the lowest in comparison to the cost of other instruments of the same type. The guitar cost the most in comparison to the cost of other instruments of the same type.

**Solution to Exercise 2.32 (p. 45)**

- a.  $IQR = 4 - 1 = 3 ; Q1 - 1.5*IQR = 1 - 1.5(3) = -3.5 ; Q3 + 1.5*IQR = 4 + 1.5(3) = 8.5$  ;The data value of 9 is larger than 8.5. The purchase of 9 books in one month is an outlier.
- b. The outlier should be investigated to see if there is an error or some other problem in the data; then a decision whether to include or exclude it should be made based on the particular situation. If it was a correct value then the data value should remain in the data set. If there is a problem with this data value, then it should be corrected or removed from the data. For example: If the data was recorded incorrectly (perhaps a 9 was miscoded and the correct value was 6) then the data should be corrected. If it was an error but the correct value is not known it should be removed from the data set.
- c.  $x\bar{ } - 2s = 2.45 - 2*1.88 = -1.31 ; x\bar{ } + 2s = 2.45 + 2*1.88 = 6.21$  ; Using this method, the five data values of 7 books purchased and the one data value of 9 books purchased would be considered unusual.
- d. No: part (a) identifies only the value of 9 to be an outlier but part (c) identifies both 7 and 9.
- e. The data is skewed (to the right). It would be more appropriate to use the method involving the IQR in part (a), identifying only the one value of 9 books purchased as an outlier. Note that part (c) remarks that identifying unusual data values by using the criteria of being further than 2 standard deviations away from the mean is most appropriate when the data are mound-shaped and symmetric.
- f. The data are skewed to the right. For skewed data it is more appropriate to use the median as a measure of center.

# Chapter 3

## Probability Topics

### 3.1 Summary of Formulas<sup>1</sup>

#### Rule 3.1: Compliment

If  $A$  and  $A'$  are complements then  $P(A) + P(A') = 1$

#### Rule 3.2: Addition Rule

$P(A \text{ OR } B) = P(A) + P(B) - P(A \text{ AND } B)$

#### Rule 3.3: Mutually Exclusive

If  $A$  and  $B$  are mutually exclusive then  $P(A \text{ AND } B) = 0$ ; so  $P(A \text{ OR } B) = P(A) + P(B)$ .

#### Rule 3.4: Multiplication Rule

- $P(A \text{ AND } B) = P(B)P(A|B)$
- $P(A \text{ AND } B) = P(A)P(B|A)$

#### Rule 3.5: Independence

If  $A$  and  $B$  are independent then:

- $P(A|B) = P(A)$
- $P(B|A) = P(B)$
- $P(A \text{ AND } B) = P(A)P(B)$

### 3.2 Homework (modified R. Bloom)<sup>2</sup>

#### Exercise 3.1

(*Solution on p. 63.*)

Suppose that you have 8 cards. 5 are green and 3 are yellow. The 5 green cards are numbered 1, 2, 3, 4, and 5. The 3 yellow cards are numbered 1, 2, and 3. The cards are well shuffled. You randomly draw one card.

- $G$  = card drawn is green
- $E$  = card drawn is even-numbered

a. List the sample space.

b.  $P(G) =$

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m16843/1.4/>>.

<sup>2</sup>This content is available online at <<http://cnx.org/content/m18924/1.2/>>.

- c.  $P(G|E) =$
- d.  $P(G \text{ AND } E) =$
- e.  $P(G \text{ OR } E) =$
- f. Are  $G$  and  $E$  mutually exclusive? Justify your answer numerically.

**Exercise 3.2**

Refer to the previous problem. Suppose that this time you randomly draw two cards, one at a time, and **with replacement**.

- $G_1$  = first card is green
  - $G_2$  = second card is green
- a. Draw a tree diagram of the situation.
  - b.  $P(G_1 \text{ AND } G_2) =$
  - c.  $P(\text{at least one green}) =$
  - d.  $P(G_2 | G_1) =$
  - e. Are  $G_2$  and  $G_1$  independent events? Explain why or why not.

**Exercise 3.3**

*(Solution on p. 63.)*

Refer to the previous problems. Suppose that this time you randomly draw two cards, one at a time, and **without replacement**.

- $G_1$  = first card is green
  - $G_2$  = second card is green
- a. Draw a tree diagram of the situation.
  - b.  $P(G_1 \text{ AND } G_2) =$
  - c.  $P(\text{at least one green}) =$
  - d.  $P(G_2 | G_1) =$
  - e. Are  $G_2$  and  $G_1$  independent events? Explain why or why not.

**Exercise 3.4**

Roll two fair dice. Each die has 6 faces.

- a. List the sample space.
- b. Let  $A$  be the event that either a 3 or 4 is rolled first, followed by an even number. Find  $P(A)$ .
- c. Let  $B$  be the event that the sum of the two rolls is at most 7. Find  $P(B)$ .
- d. In words, explain what " $P(A|B)$ " represents. Find  $P(A|B)$ .
- e. Are  $A$  and  $B$  mutually exclusive events? Explain your answer in 1 - 3 complete sentences, including numerical justification.
- f. Are  $A$  and  $B$  independent events? Explain your answer in 1 - 3 complete sentences, including numerical justification.

**Exercise 3.5**

*(Solution on p. 63.)*

A special deck of cards has 10 cards. Four are green, three are blue, and three are red. When a card is picked, the color of it is recorded. An experiment consists of first picking a card and then tossing a coin.

- a. List the sample space.

- b. Let  $A$  be the event that a blue card is picked first, followed by landing a head on the coin toss. Find  $P(A)$ .
- c. Let  $B$  be the event that a red or green is picked, followed by landing a head on the coin toss. Are the events  $A$  and  $B$  mutually exclusive? Explain your answer in 1 - 3 complete sentences, including numerical justification.
- d. Let  $C$  be the event that a red or blue is picked, followed by landing a head on the coin toss. Are the events  $A$  and  $C$  mutually exclusive? Explain your answer in 1 - 3 complete sentences, including numerical justification.

**Exercise 3.6**

An experiment consists of first rolling a die and then tossing a coin:

- a. List the sample space.
- b. Let  $A$  be the event that either a 3 or 4 is rolled first, followed by landing a head on the coin toss. Find  $P(A)$ .
- c. Let  $B$  be the event that a number less than 2 is rolled, followed by landing a head on the coin toss. Are the events  $A$  and  $B$  mutually exclusive? Explain your answer in 1 - 3 complete sentences, including numerical justification.

**Exercise 3.7**

*(Solution on p. 63.)*

An experiment consists of tossing a nickel, a dime and a quarter. Of interest is the side the coin lands on.

- a. List the sample space.
- b. Let  $A$  be the event that there are at least two tails. Find  $P(A)$ .
- c. Let  $B$  be the event that the first and second tosses land on heads. Are the events  $A$  and  $B$  mutually exclusive? Explain your answer in 1 - 3 complete sentences, including justification.

**Exercise 3.8**

Consider the following scenario:

- Let  $P(C) = 0.4$
- Let  $P(D) = 0.5$
- Let  $P(C \mid D) = 0.6$

- a. Find  $P(C \text{ AND } D)$ .
- b. Are  $C$  and  $D$  mutually exclusive? Why or why not?
- c. Are  $C$  and  $D$  independent events? Why or why not?
- d. Find  $P(C \text{ OR } D)$ .
- e. Find  $P(D \mid C)$ .

**Exercise 3.9**

*(Solution on p. 63.)*

$E$  and  $F$  mutually exclusive events.  $P(E) = 0.4$ ;  $P(F) = 0.5$ . Find  $P(E \mid F)$ .

**Exercise 3.10**

$J$  and  $K$  are independent events.  $P(J \mid K) = 0.3$ . Find  $P(J)$ .

**Exercise 3.11**

*(Solution on p. 63.)*

$U$  and  $V$  are mutually exclusive events.  $P(U) = 0.26$ ;  $P(V) = 0.37$ . Find:

- a.  $P(U \text{ AND } V) =$
- b.  $P(U \mid V) =$
- c.  $P(U \text{ OR } V) =$

**Exercise 3.12**

$Q$  and  $R$  are independent events.  $P(Q) = 0.4$ ;  $P(Q \text{ AND } R) = 0.1$ . Find  $P(R)$ .

**Exercise 3.13**

$Y$  and  $Z$  are independent events.

(*Solution on p. 63.*)

- a. Rewrite the basic Addition Rule  $P(Y \text{ OR } Z) = P(Y) + P(Z) - P(Y \text{ AND } Z)$  using the information that  $Y$  and  $Z$  are independent events.
- b. Use the rewritten rule to find  $P(Z)$  if  $P(Y \text{ OR } Z) = 0.71$  and  $P(Y) = 0.42$ .

**Exercise 3.14**

$G$  and  $H$  are mutually exclusive events.  $P(G) = 0.5$ ;  $P(H) = 0.3$

- a. Explain why the following statement MUST be false:  $P(H | G) = 0.4$ .
- b. Find:  $P(H \text{ OR } G)$ .
- c. Are  $G$  and  $H$  independent or dependent events? Explain in a complete sentence.

**Exercise 3.15**

(*Solution on p. 63.*)

The following are real data from Santa Clara County, CA. As of March 31, 2000, there was a total of 3059 documented cases of AIDS in the county. They were grouped into the following categories (Source: *Santa Clara County Public H.D.*):

	Homosexual/Bisexual	IV Drug User*	Heterosexual Contact	Other	Totals
Female	0	70	136	49	—
Male	2146	463	60	135	—
Totals	—	—	—	—	—

Table 3.1: \* includes homosexual/bisexual IV drug users

Suppose one of the persons with AIDS in Santa Clara County is randomly selected. Compute the following:

- a.  $P(\text{person is female}) =$
- b.  $P(\text{person has a risk factor Heterosexual Contact}) =$
- c.  $P(\text{person is female OR has a risk factor of IV Drug User}) =$
- d.  $P(\text{person is female AND has a risk factor of Homosexual/Bisexual}) =$
- e.  $P(\text{person is male AND has a risk factor of IV Drug User}) =$
- f.  $P(\text{female GIVEN person got the disease from heterosexual contact}) =$
- g. Construct a Venn Diagram. Make one group females and the other group heterosexual contact.

**Exercise 3.16**

Solve these questions using probability rules. Do NOT use the contingency table above. 3059 cases of AIDS had been reported in Santa Clara County, CA, through March 31, 2000. Those cases will be our population. Of those cases, 6.4% obtained the disease through heterosexual contact and 7.4% are female. Out of the females with the disease, 53.3% got the disease from heterosexual contact.

- a.  $P(\text{person is female}) =$
- b.  $P(\text{person obtained the disease through heterosexual contact}) =$
- c.  $P(\text{female GIVEN person got the disease from heterosexual contact}) =$

- d. Construct a Venn Diagram. Make one group females and the other group heterosexual contact. Fill in all values as probabilities.

**Exercise 3.17**

(*Solution on p. 64.*)

The following table identifies a group of children by one of four hair colors, and by type of hair.

Hair Type	Brown	Blond	Black	Red	Totals
Wavy	20		15	3	43
Straight	80	15		12	
Totals		20			215

Table 3.2

- a. Complete the table above.
- b. What is the probability that a randomly selected child will have wavy hair?
- c. What is the probability that a randomly selected child will have either brown or blond hair?
- d. What is the probability that a randomly selected child will have wavy brown hair?
- e. What is the probability that a randomly selected child will have red hair, given that he has straight hair?
- f. If B is the event of a child having brown hair, find the probability of the complement of B.
- g. In words, what does the complement of B represent?

**Exercise 3.18**

A previous year, the weights of the members of the **San Francisco 49ers** and the **Dallas Cowboys** were published in the *San Jose Mercury News*. The factual data are compiled into the following table.

Shirt#	$\leq 210$	211-250	251-290	$290 \leq$
1-33	21	5	0	0
34-66	6	18	7	4
66-99	6	12	22	5

Table 3.3

For the following, suppose that you randomly select one player from the 49ers or Cowboys.

- a. Find the probability that his shirt number is from 1 to 33.
- b. Find the probability that he weighs at most 210 pounds.
- c. Find the probability that his shirt number is from 1 to 33 AND he weighs at most 210 pounds.
- d. Find the probability that his shirt number is from 1 to 33 OR he weighs at most 210 pounds.
- e. Find the probability that his shirt number is from 1 to 33 GIVEN that he weighs at most 210 pounds.
- f. If having a shirt number from 1 to 33 and weighing at most 210 pounds were independent events, then what should be true about  $P(\text{Shirt\# } 1-33 \mid \leq 210 \text{ pounds})$ ?

**Exercise 3.19**

(Solution on p. 64.)

Approximately 249,000,000 people live in the United States. Of these people, 31,800,000 speak a language other than English at home. Of those who speak another language at home, over 50 percent speak Spanish. (Source: U.S. Bureau of the Census, 1990 Census)

Let:  $E$  = speak English at home;  $E'$  = speak another language at home;  $S$  = speak Spanish at home

Finish each probability statement by matching the correct answer.

Probability Statements	Answers
a. $P(E') =$	i. 0.8723
b. $P(E) =$	ii. $> 0.50$
c. $P(S) =$	iii. 0.1277
d. $P(S   E') =$	iv. $> 0.0639$

Table 3.4

**Exercise 3.20**

The probability that a male develops some form of cancer in his lifetime is 0.4567 (Source: American Cancer Society). The probability that a male has at least one false positive test result (meaning the test comes back for cancer when the man does not have it) is 0.51 (Source: USA Today). Some of the questions below do not have enough information for you to answer them. Write “not enough information” for those answers.

Let:  $C$  = a man develops cancer in his lifetime;  $P$  = man has at least one false positive

- a. Construct a tree diagram of the situation.
- b.  $P(C) =$
- c.  $P(P|C) =$
- d.  $P(P|C') =$
- e. If a test comes up positive, based upon numerical values, can you assume that man has cancer? Justify numerically and explain why or why not.

**Exercise 3.21**

(Solution on p. 64.)

In 1994, the U.S. government held a lottery to issue 55,000 Green Cards (permits for non-citizens to work legally in the U.S.). Renate Deutsch, from Germany, was one of approximately 6.5 million people who entered this lottery. Let  $G$  = won Green Card.

- a. What was Renate’s chance of winning a Green Card? Write your answer as a probability statement.
- b. In the summer of 1994, Renate received a letter stating she was one of 110,000 finalists chosen. Once the finalists were chosen, assuming that each finalist had an equal chance to win, what was Renate’s chance of winning a Green Card? Let  $F$  = was a finalist. Write your answer as a conditional probability statement.
- c. Are  $G$  and  $F$  independent or dependent events? Justify your answer numerically and also explain why.
- d. Are  $G$  and  $F$  mutually exclusive events? Justify your answer numerically and also explain why.

NOTE: P.S. Amazingly, on 2/1/95, Renate learned that she would receive her Green Card – true story!

**Exercise 3.22**

Three professors at George Washington University did an experiment to determine if economists are more selfish than other people. They dropped 64 stamped, addressed envelopes with \$10 cash in different classrooms on the George Washington campus. 44% were returned overall. From the economics classes 56% of the envelopes were returned. From the business, psychology, and history classes 31% were returned. (Source: *Wall Street Journal*)

Let:  $R$  = money returned;  $E$  = economics classes;  $O$  = other classes

- Write a probability statement for the overall percent of money returned.
- Write a probability statement for the percent of money returned out of the economics classes.
- Write a probability statement for the percent of money returned out of the other classes.
- Is money being returned independent of the class? Justify your answer numerically and explain it.
- Based upon this study, do you think that economists are more selfish than other people? Explain why or why not. Include numbers to justify your answer.

**Exercise 3.23**

(*Solution on p. 64.*)

The chart below gives the number of suicides estimated in the U.S. for a recent year by age, race (black and white), and sex. We are interested in possible relationships between age, race, and sex. We will let suicide victims be our population. (Source: *The National Center for Health Statistics, U.S. Dept. of Health and Human Services*)

Race and Sex	1 - 14	15 - 24	25 - 64	over 64	TOTALS
white, male	210	3360	13,610		22,050
white, female	80	580	3380		4930
black, male	10	460	1060		1670
black, female	0	40	270		330
all others					
TOTALS	310	4650	18,780		29,760

Table 3.5

NOTE: Do not include "all others" for parts (f), (g), and (i).

- Fill in the column for the suicides for individuals over age 64.
- Fill in the row for all other races.
- Find the probability that a randomly selected individual was a white male.
- Find the probability that a randomly selected individual was a black female.
- Find the probability that a randomly selected individual was black
- Comparing "Race and Sex" to "Age," which two groups are mutually exclusive? How do you know?
- Find the probability that a randomly selected individual was male.
- Out of the individuals over age 64, find the probability that a randomly selected individual was a black or white male.
- Are being male and committing suicide over age 64 independent events? How do you know?

**The next two questions refer to the following:** The percent of licensed U.S. drivers (from a recent year) that are female is 48.60. Of the females, 5.03% are age 19 and under; 81.36% are age 20 - 64; 13.61% are age 65 or over. Of the licensed U.S. male drivers, 5.04% are age 19 and under; 81.43% are age 20 - 64; 13.53% are age 65 or over. (Source: Federal Highway Administration, U.S. Dept. of Transportation)

#### Exercise 3.24

Complete the following:

- a. Construct a table or a tree diagram of the situation.
- b.  $P(\text{driver is female}) =$
- c.  $P(\text{driver is age 65 or over} \mid \text{driver is female}) =$
- d.  $P(\text{driver is age 65 or over AND female}) =$
- e. In words, explain the difference between the probabilities in part (c) and part (d).
- f.  $P(\text{driver is age 65 or over}) =$
- g. Are being age 65 or over and being female mutually exclusive events? How do you know?
- h.  $P(\text{driver is "male"} \text{ OR } \text{"age 19 or under"}) =$

#### Exercise 3.25

(*Solution on p. 64.*)

Suppose that 10,000 U.S. licensed drivers are randomly selected.

- a. How many would you expect to be male?
- b. Using the table or tree diagram from the previous exercise, construct a contingency table of gender versus age group.
- c. Using the contingency table, find the probability that out of the age 20 - 64 group, a randomly selected driver is female.

#### Exercise 3.26

Approximately 86.5% of Americans commute to work by car, truck or van. Out of that group, 84.6% drive alone and 15.4% drive in a carpool. Approximately 3.9% walk to work and approximately 5.3% take public transportation. (*Source: Bureau of the Census, U.S. Dept. of Commerce. Disregard rounding approximations.*)

- a. Construct a table or a tree diagram of the situation. Include a branch for all other modes of transportation to work.
- b. Assuming that the walkers walk alone, what percent of all commuters travel alone to work?
- c. Suppose that 1000 workers are randomly selected. How many would you expect to travel alone to work?
- d. Suppose that 1000 workers are randomly selected. How many would you expect to drive in a carpool?
- e. What percent of workers do NOT "drive alone"?

#### Exercise 3.27

Explain what is wrong with the following statements. Use complete sentences.

- a. If there's a 60% chance of rain on Saturday and a 70% chance of rain on Sunday, then there's a 130% chance of rain over the weekend.
- b. The probability that a baseball player hits a home run is greater than the probability that he gets a successful hit.

### 3.2.1 Questions 28 through 32 are multiple choice

Questions 28 and 29 refer to the following probability tree diagram which shows tossing an unfair coin FOLLOWED BY drawing one bead from a cup containing 3 red ( $R$ ), 4 yellow ( $Y$ ) and 5 blue ( $B$ ) beads. For the coin,  $P(H) = \frac{2}{3}$  and  $P(T) = \frac{1}{3}$  where  $H$  = "heads" and  $T$  = "tails".

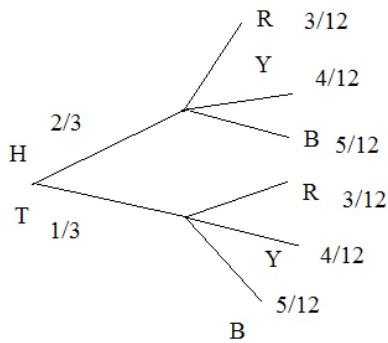


Figure 3.1

#### Exercise 3.28

Find  $P(\text{tossing a Head on the coin AND a Red bead})$

(Solution on p. 64.)

- A.  $\frac{2}{3}$
- B.  $\frac{5}{15}$
- C.  $\frac{6}{36}$
- D.  $\frac{5}{36}$

#### Exercise 3.29

Find  $P(\text{Blue bead})$ .

(Solution on p. 64.)

- A.  $\frac{15}{36}$
- B.  $\frac{10}{36}$
- C.  $\frac{10}{12}$
- D.  $\frac{6}{36}$

Questions 30 through 32 refer to the following table of data obtained from [www.baseball-almanac.com](http://www.baseball-almanac.com)<sup>3</sup> showing hit information for 4 well known baseball players.

<sup>3</sup><http://cnx.org/content/m18924/latest/www.baseball-almanac.com>

NAME	Single	Double	Triple	Home Run	TOTAL HITS
Babe Ruth	1517	506	136	714	2873
Jackie Robinson	1054	273	54	137	1518
Ty Cobb	3603	174	295	114	4189
Hank Aaron	2294	624	98	755	3771
TOTAL	8471	1577	583	1720	12351

Table 3.6

**Exercise 3.30**

(Solution on p. 64.)

Find  $P(\text{hit was made by Babe Ruth})$ .

- A.  $\frac{1518}{2873}$
- B.  $\frac{2873}{12351}$
- C.  $\frac{583}{12351}$
- D.  $\frac{4189}{12351}$

**Exercise 3.31**

(Solution on p. 64.)

Find  $P(\text{hit was made by Ty Cobb} \mid \text{The hit was a Home Run})$ 

- A.  $\frac{4189}{12351}$
- B.  $\frac{1141}{1720}$
- C.  $\frac{1720}{4189}$
- D.  $\frac{114}{12351}$

**Exercise 3.32**

(Solution on p. 64.)

Are the hit being made by Hank Aaron and the hit being a double independent events?

- A. Yes, because  $P(\text{hit by Hank Aaron} \mid \text{hit is a double}) = P(\text{hit by Hank Aaron})$
- B. No, because  $P(\text{hit by Hank Aaron} \mid \text{hit is a double}) \neq P(\text{hit is a double})$
- C. No, because  $P(\text{hit is by Hank Aaron} \mid \text{hit is a double}) \neq P(\text{hit by Hank Aaron})$
- D. Yes, because  $P(\text{hit is by Hank Aaron} \mid \text{hit is a double}) = P(\text{hit is a double})$

**Exercise 3.33**

(Solution on p. 64.)

Given events G and H:  $P(G) = 0.43$ ;  $P(H) = 0.26$ ;  $P(H \text{ and } G) = 0.14$ 

- a. Find  $P(H \text{ or } G)$
- b. Find the probability of the complement of event (H and G)
- c. Find the probability of the complement of event (H or G)

**Exercise 3.34**

(Solution on p. 65.)

Given events J and K:  $P(J) = 0.18$ ;  $P(K) = 0.37$ ;  $P(J \text{ or } K) = 0.45$ 

- a. Find  $P(J \text{ and } K)$
- b. Find the probability of the complement of event (J and K)
- c. Find the probability of the complement of event (J or K)

**Exercise 3.35***(Solution on p. 65.)*

United Blood Services is a blood bank that serves more than 500 hospitals in 18 states. According to their website, <http://www.unitedbloodservices.org/humanbloodtypes.html>, a person with type O blood and a negative Rh factor ( $Rh-$ ) can donate blood to any person with any bloodtype. Their data show that 43% of people have type O blood and 15% of people have  $Rh-$  factor; 52% of people have type O or  $Rh-$  factor.

- a. Find the probability that a person has both type O blood and the  $Rh-$  factor
- b. Find the probability that a person does NOT have both type O blood and the  $Rh-$  factor.

**Exercise 3.36***(Solution on p. 65.)*

At a college, 72% of courses have final exams and 46% of courses require research papers. Suppose that 32% of courses have a research paper and a final exam. Let  $F$  be the event that a course has a final exam. Let  $R$  be the event that a course requires a research paper.

- a. Find the probability that a course has a final exam or a research project.
- b. Find the probability that a course has NEITHER of these two requirements.

**Exercise 3.37***(Solution on p. 65.)*

In a box of assorted cookies, 36% contain chocolate and 12% contain nuts. Of those, 8% contain both chocolate and nuts. Sean is allergic to both chocolate and nuts.

- a. Find the probability that a cookie contains chocolate or nuts (he can't eat it).
- b. Find the probability that a cookie does not contain chocolate or nuts (he can eat it).

**Exercise 3.38***(Solution on p. 65.)*

A college finds that 10% of students have taken a distance learning class and that 40% of students are part time students. Of the part time students, 20% have taken a distance learning class. Let  $D$  = event that a student takes a distance learning class and  $E$  = event that a student is a part time student

- a. Find  $P(D \text{ and } E)$
- b. Find  $P(E | D)$
- c. Find  $P(D \text{ or } E)$
- d. Using an appropriate test, show whether  $D$  and  $E$  are independent.
- e. Using an appropriate test, show whether  $D$  and  $E$  are mutually exclusive.

**Exercise 3.39***(Solution on p. 65.)*

At a certain store the manager has determined that 30% of customers pay cash and 70% of customers pay by debit card. (No other method of payment is accepted.) Let  $M$  = event that a customer pays cash and  $D$  = event that a customer pays by debit card.

- a. Suppose two customers (Al and Betty) come to the store. Explain why it would be reasonable to assume that their choices of payment methods are independent of each other.
- b. Draw the tree that represents the all possibilities for the 2 customers and their methods of payment. Write the probabilities along each branch of the tree.
- c. For each complete path through the tree, write the event it represents and find the probability.
- d. Let  $S$  be the event that both customers use the same method of payment. Find  $P(S)$
- e. Let  $T$  be the event that both customers use different methods of payment. Find  $P(T)$  by two different methods: by using the complement rule and by using the branches of the tree. Your answers should be the same with both methods.
- f. Let  $U$  be the event that the second customer uses a debit card. Find  $P(U)$

**Exercise 3.40**

(Solution on p. 65.)

A box of cookies contains 3 chocolate and 7 butter cookies. Miguel randomly selects a cookie and eats it. Then he randomly selects another cookie and eats it also. (How many cookies did he take?)

- a. Are the probabilities for the flavor of the SECOND cookie that Miguel selects independent of his first selection, or do the probabilities depend on the type of cookie that Miguel selected first? Explain.
- b. Draw the tree that represents the possibilities for the cookie selections. Write the probabilities along each branch of the tree.
- c. For each complete path through the tree, write the event it represents and find the probabilities.
- d. Let  $S$  be the event that both cookies selected were the same flavor. Find  $P(S)$ .
- e. Let  $T$  be the event that both cookies selected were different flavors. Find  $P(T)$  by two different methods: by using the complement rule and by using the branches of the tree. Your answers should be the same with both methods.
- f. Let  $U$  be the event that the second cookie selected is a butter cookie. Find  $P(U)$ .

**Exercise 3.41**

(Solution on p. 65.)

When the Euro coin was introduced in 2002, two math professors had their statistics students test whether the Belgian 1 Euro coin was a fair coin. They spun the coin rather than tossing it, and it was found that out of 250 spins, 140 showed a head (event  $H$ ) while 110 showed a tail (event  $T$ ). Therefore, they claim that this is not a fair coin.

- a. Based on the data above, find  $P(H)$  and  $P(T)$ .
- b. Use a tree to find the probabilities of each possible outcome for the experiment of tossing the coin twice.
- c. Use the tree to find the probability of obtaining exactly one head in two tosses of the coin.
- d. Use the tree to find the probability of obtaining at least one head.

### 3.3 Review Questions<sup>4</sup>

**The first six exercises refer to the following study:** In a survey of 100 stocks on NASDAQ, the average percent increase for the past year was 9% for NASDAQ stocks. Answer the following:

**Exercise 3.42: REVIEW QUESTION 1**

(Solution on p. 65.)

The “average increase” for all NASDAQ stocks is the:

- A. Population
- B. Statistic
- C. Parameter
- D. Sample
- E. Variable

**Exercise 3.43: REVIEW QUESTION 2**

(Solution on p. 65.)

All of the NASDAQ stocks are the:

- A. Population
- B. Statistic
- C. Parameter

---

<sup>4</sup>This content is available online at <<http://cnx.org/content/m19023/1.1/>>.

- D. Sample
- E. Variable

**Exercise 3.44: REVIEW QUESTION 3***(Solution on p. 65.)*

9% is the:

- A. Population
- B. Statistic
- C. Parameter
- D. Sample
- E. Variable

**Exercise 3.45: REVIEW QUESTION 4***(Solution on p. 65.)*

The 100 NASDAQ stocks in the survey are the:

- A. Population
- B. Statistic
- C. Parameter
- D. Sample
- E. Variable

**Exercise 3.46: REVIEW QUESTION 5***(Solution on p. 65.)*

The percent increase for one stock in the survey is the:

- A. Population
- B. Statistic
- C. Parameter
- D. Sample
- E. Variable

**Exercise 3.47: REVIEW QUESTION 6***(Solution on p. 65.)*

Would the data collected be qualitative, quantitative – discrete, or quantitative – continuous?

**The next two questions refer to the following study:** Thirty people spent two weeks around Mardi Gras in New Orleans. Their two-week weight gain is below. (Note: a loss is shown by a negative weight gain.)

Weight Gain	Frequency
-2	3
-1	5
0	2
1	4
4	13
6	2
11	1

**Table 3.7****Exercise 3.48: REVIEW QUESTION 7***(Solution on p. 66.)*

Calculate the following values:

- a. The average weight gain for the two weeks
- b. The standard deviation
- c. The first, second, and third quartiles

**Exercise 3.49: REVIEW QUESTION 8**

Construct a histogram and a boxplot of the data.

## Solutions to Exercises in Chapter 3

### Solution to Exercise 3.1 (p. 49)

- a.  $\{G1, G2, G3, G4, G5, Y1, Y2, Y3\}$
- b.  $\frac{5}{8}$
- c.  $\frac{2}{3}$
- d.  $\frac{2}{8}$
- e.  $\frac{6}{8}$
- f. No

### Solution to Exercise 3.3 (p. 50)

- b.  $\left(\frac{5}{8}\right) \left(\frac{4}{7}\right)$
- c.  $\left(\frac{5}{8}\right) \left(\frac{3}{7}\right) + \left(\frac{3}{8}\right) \left(\frac{5}{7}\right) + \left(\frac{5}{8}\right) \left(\frac{4}{7}\right)$
- d.  $\frac{4}{7}$
- e. No

### Solution to Exercise 3.5 (p. 50)

- a. {GH, GT, BH, BT, RH, RT}
- b.  $\frac{3}{20}$
- c. Yes
- d. No

### Solution to Exercise 3.7 (p. 51)

- a.  $\{(HHH), (HHT), (HTH), (HTT), (THH), (THT), (TTH), (TTT)\}$
- b.  $\frac{4}{8}$
- c. Yes

### Solution to Exercise 3.9 (p. 51)

0

### Solution to Exercise 3.11 (p. 51)

- a. 0
- b. 0
- c. 0.63

### Solution to Exercise 3.13 (p. 52)

- b. 0.5

### Solution to Exercise 3.15 (p. 52)

The completed contingency table is as follows:

	Homosexual/Bisexual	IV Drug User*	Heterosexual Contact	Other	Totals
Female	0	70	136	49	255
Male	2146	463	60	135	2804
Totals	2146	533	196	174	3059

Table 3.8: \* includes homosexual/bisexual IV drug users

- a.  $\frac{255}{3059}$
- b.  $\frac{196}{3059}$
- c.  $\frac{718}{3059}$
- d. 0
- e.  $\frac{463}{3059}$
- f.  $\frac{136}{196}$

**Solution to Exercise 3.17 (p. 53)**

- b.  $\frac{43}{215}$
- c.  $\frac{120}{215}$
- d.  $\frac{20}{215}$
- e.  $\frac{12}{172}$
- f.  $\frac{115}{215}$

**Solution to Exercise 3.19 (p. 54)**

- a. iii
- b. i
- c. iv
- d. ii

**Solution to Exercise 3.21 (p. 54)**

- a.  $P(G) = 0.008$
- b. 0.5
- c. dependent
- d. No

**Solution to Exercise 3.23 (p. 55)**

- c.  $\frac{22050}{29760}$
- d.  $\frac{330}{29760}$
- e.  $\frac{2000}{29760}$
- f.  $\frac{23720}{29760}$
- g.  $\frac{5010}{6020}$
- h. Black females and ages 1-14
- i. No

**Solution to Exercise 3.25 (p. 56)**

- a. 5140
- c. 0.49

**Solution to Exercise 3.28 (p. 57)**

C

**Solution to Exercise 3.29 (p. 57)**

A

**Solution to Exercise 3.30 (p. 58)**

B

**Solution to Exercise 3.31 (p. 58)**

B

**Solution to Exercise 3.32 (p. 58)**

C

**Solution to Exercise 3.33 (p. 58)**

- a.  $P(H \text{ or } G) = P(H) + P(G) - P(H \text{ and } G) = 0.26 + 0.43 - 0.14 = 0.55$
- b.  $P(\text{NOT}(H \text{ and } G)) = 1 - P(H \text{ and } G) = 1 - 0.14 = 0.86$
- c.  $P(\text{NOT}(H \text{ or } G)) = 1 - P(H \text{ or } G) = 1 - 0.55 = 0.45$

**Solution to Exercise 3.34 (p. 58)**

- a.  $P(J \text{ or } K) = P(J) + P(K) - P(J \text{ and } K); 0.45 = 0.18 + 0.37 - P(J \text{ and } K); \text{ solve to find } P(J \text{ and } K) = 0.10$
- b.  $P(\text{NOT}(J \text{ and } K)) = 1 - P(J \text{ and } K) = 1 - 0.10 = 0.90$
- c.  $P(\text{NOT}(J \text{ or } K)) = 1 - P(J \text{ or } K) = 1 - 0.45 = 0.55$

**Solution to Exercise 3.35 (p. 59)**

- a.  $P(\text{Type O or Rh-}) = P(\text{Type O}) + P(\text{Rh-}) - P(\text{Type O and Rh-})$
- a.  $0.52 = 0.43 + 0.15 - P(\text{Type O and Rh-}); \text{ solve to find } P(\text{Type O and Rh-}) = 0.06$
- a. 6% of people have type O Rh- blood
- b.  $P(\text{NOT}(\text{Type O and Rh-})) = 1 - P(\text{Type O and Rh-}) = 1 - 0.06 = 0.94$
- b. 94% of people do not have type O Rh- blood

**Solution to Exercise 3.36 (p. 59)**

- a.  $P(R \text{ or } F) = P(R) + P(F) - P(R \text{ and } F) = 0.72 + 0.46 - 0.32 = 0.86$
- b.  $P(\text{Neither R nor F}) = 1 - P(R \text{ or } F) = 1 - 0.86 = 0.14$

**Solution to Exercise 3.37 (p. 59)**

- . Let C be the event that the cookie contains chocolate. Let N be the event that the cookie contains nuts.
- a.  $P(C \text{ or } N) = P(C) + P(N) - P(C \text{ and } N) = 0.36 + 0.12 - 0.08 = 0.40$
- b.  $P(\text{neither chocolate nor nuts}) = 1 - P(C \text{ or } N) = 1 - 0.60 = 0.40$

**Solution to Exercise 3.38 (p. 59)**

- a.  $P(D \text{ and } E) = P(D | E)P(E) = (0.20)(0.40) = 0.08$
- b.  $P(E | D) = P(D \text{ and } E) / P(D) = 0.08 / 0.10 = 0.80$
- c.  $P(D \text{ or } E) = P(D) + P(E) - P(D \text{ and } E) = 0.10 + 0.40 - 0.08 = 0.42$
- d. Not Independent:  $P(D | E) = 0.20$  which does not equal  $P(D) = .10$
- e. Not Mutually Exclusive:  $P(D \text{ and } E) = 0.08$ ; if they were mutually exclusive then we would need to have  $P(D \text{ and } E) = 0$ , which is not true here.

**Solution to Exercise 3.39 (p. 59)**

Solution is posted on instructor's website for this class.

**Solution to Exercise 3.40 (p. 60)**

Solution is posted on instructor's website for this class.

**Solution to Exercise 3.41 (p. 60)**

Solution is posted on instructor's website for this class.

**Solution to Exercise 3.42 (p. 60)**

REVIEW QUESTION 1 Solution: C. Parameter

**Solution to Exercise 3.43 (p. 60)**

REVIEW QUESTION 2 Solution : A. Population

**Solution to Exercise 3.44 (p. 61)**

REVIEW QUESTION 3 Solution : B. Statistic

**Solution to Exercise 3.45 (p. 61)**

REVIEW QUESTION 4 Solution : D. Sample

**Solution to Exercise 3.46 (p. 61)**

REVIEW QUESTION 5 Solution : E. Variable

**Solution to Exercise 3.47 (p. 61)**

REVIEW QUESTION 6 Solution: quantitative - continuous

**Solution to Exercise 3.48 (p. 61)**

REVIEW QUESTION 7 Solution

- a. 2.27
- b. 3.04
- c. -1, 4, 4

## Chapter 4

# Discrete Random Variables

### 4.1 Summary of Functions<sup>1</sup>

**Formula 4.1:** Binomial

$$X \sim B(n, p)$$

$X$  = the number of successes in  $n$  independent trials

$n$  = the number of independent trials

$X$  takes on the values  $x = 0, 1, 2, 3, \dots, n$

$p$  = the probability of a success for any trial

$q$  = the probability of a failure for any trial

$$p + q = 1 \quad q = 1 - p$$

The mean is  $\mu = np$ . The standard deviation is  $\sigma = \sqrt{npq}$ .

**Formula 4.2:** Geometric

$$X \sim G(p)$$

$X$  = the number of independent trials until the first success (count the failures and the first success)

$X$  takes on the values  $x = 1, 2, 3, \dots$

$p$  = the probability of a success for any trial

$q$  = the probability of a failure for any trial

$$p + q = 1$$

$$q = 1 - p$$

$$\text{The mean is } \mu = \frac{1}{p}$$

$$\text{The standard deviation is } \sigma = \sqrt{\frac{1}{p} \left( \left( \frac{1}{p} \right) - 1 \right)}$$

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m16833/1.8/>>.

**Formula 4.3:** Hypergeometric  
 $X \sim H(r, b, n)$

$X$  = the number of items from the group of interest that are in the chosen sample.

$X$  may take on the values  $x = 0, 1, \dots$ , up to the size of the group of interest. (The minimum value for  $X$  may be larger than 0 in some instances.)

$r$  = the size of the group of interest (first group)

$b$  = the size of the second group

$n$  = the size of the chosen sample.

$$n \leq r + b$$

The mean is:  $\mu = \frac{nr}{r+b}$

The standard deviation is:  $\sigma = \sqrt{\frac{rbn(r+b+n)}{(r+b)^2(r+b-1)}}$

**Formula 4.4:** Poisson

$X \sim P(\mu)$

$X$  = the number of occurrences in the interval of interest

$X$  takes on the values  $x = 0, 1, 2, 3, \dots$

The mean  $\mu$  is typically given. ( $\lambda$  is often used as the mean instead of  $\mu$ .) When the Poisson is used to approximate the binomial, we use the binomial mean  $\mu = np$ .  $n$  is the binomial number of trials.  $p$  = the probability of a success for each trial. This formula is valid when  $n$  is "large" and  $p$  "small" (a general rule is that  $n$  should be greater than or equal to 20 and  $p$  should be less than or equal to 0.05). If  $n$  is large enough and  $p$  is small enough then the Poisson approximates the binomial very well. The standard deviation is  $\sigma = \sqrt{\mu}$ .

## 4.2 Homework (modified R. Bloom)<sup>2</sup>

### Exercise 4.1

(Solution on p. 82.)

1. Complete the PDF and answer the questions.

$X$	$P(X = x)$	$X \cdot P(X = x)$
0	0.3	
1	0.2	
2		
3	0.4	

Table 4.1

- a. Find the probability that  $X = 2$ .

<sup>2</sup>This content is available online at <<http://cnx.org/content/m18927/1.2/>>.

- b. Find the expected value.

**Exercise 4.2**

Suppose that you are offered the following “deal.” You roll a die. If you roll a 6, you win \$10. If you roll a 4 or 5, you win \$5. If you roll a 1, 2, or 3, you pay \$6.

- a. What are you ultimately interested in here (the value of the roll or the money you win)?
- b. In words, define the Random Variable  $X$ .
- c. List the values that  $X$  may take on.
- d. Construct a PDF.
- e. Over the long run of playing this game, what are your expected average winnings per game?
- f. Based on numerical values, should you take the deal? Explain your decision in complete sentences.

**Exercise 4.3**

(*Solution on p. 82.*)

A venture capitalist, willing to invest \$1,000,000, has three investments to choose from. The first investment, a software company, has a 10% chance of returning \$5,000,000 profit, a 30% chance of returning \$1,000,000 profit, and a 60% chance of losing the million dollars. The second company, a hardware company, has a 20% chance of returning \$3,000,000 profit, a 40% chance of returning \$1,000,000 profit, and a 40% chance of losing the million dollars. The third company, a biotech firm, has a 10% chance of returning \$6,000,000 profit, a 70% of no profit or loss, and a 20% chance of losing the million dollars.

- a. Construct a PDF for each investment.
- b. Find the expected value for each investment.
- c. Which is the safest investment? Why do you think so?
- d. Which is the riskiest investment? Why do you think so?
- e. Which investment has the highest expected return, on average?

**Exercise 4.4**

A theater group holds a fund-raiser. It sells 100 raffle tickets for \$5 apiece. Suppose you purchase 4 tickets. The prize is 2 passes to a Broadway show, worth a total of \$150.

- a. What are you interested in here?
- b. In words, define the Random Variable  $X$ .
- c. List the values that  $X$  may take on.
- d. Construct a PDF.
- e. If this fund-raiser is repeated often and you always purchase 4 tickets, what would be your expected average winnings per game?

**Exercise 4.5**

(*Solution on p. 82.*)

Suppose that 20,000 married adults in the United States were randomly surveyed as to the number of children they have. The results are compiled and are used as theoretical probabilities. Let  $X$  = the number of children

$X$	$P(X = x)$	$X \cdot P(X = x)$
0	0.10	
1	0.20	
2	0.30	
3		
4	0.10	
5	0.05	
6 (or more)	0.05	

**Table 4.2**

- a. Find the probability that a married adult has 3 children.
- b. In words, what does the expected value in this example represent?
- c. Find the expected value.
- d. Is it more likely that a married adult will have 2 – 3 children or 4 – 6 children? How do you know?

**Exercise 4.6**

Suppose that the PDF for the number of years it takes to earn a Bachelor of Science (B.S.) degree is given below.

$X$	$P(X = x)$
3	0.05
4	0.40
5	0.30
6	0.15
7	0.10

**Table 4.3**

- a. In words, define the Random Variable  $X$ .
- b. What does it mean that the values 0, 1, and 2 are not included for  $X$  on the PDF?
- c. On average, how many years do you expect it to take for an individual to earn a B.S.?

**4.2.1 For each problem:**

- a. In words, define the Random Variable  $X$ .
- b. List the values that  $X$  may take on.
- c. Give the distribution of  $X$ .  $X \sim$

Then, answer the questions specific to each individual problem.

**Exercise 4.7**

(*Solution on p. 82.*)

Six different colored dice are rolled. Of interest is the number of dice that show a “1.”

- d. On average, how many dice would you expect to show a “1”?
- e. Find the probability that all six dice show a “1.”
- f. Is it more likely that 3 or that 4 dice will show a “1”? Use numbers to justify your answer numerically.

**Exercise 4.8**

According to a 2003 publication by Waits and Lewis (source: [http://nces.ed.gov/pubs2003/2003017.pdf<sup>3</sup>](http://nces.ed.gov/pubs2003/2003017.pdf)), by the end of 2002, 92% of U.S. public two-year colleges offered distance learning courses. Suppose you randomly pick 13 U.S. public two-year colleges. We are interested in the number that offer distance learning courses.

- d. On average, how many schools would you expect to offer such courses?
- e. Find the probability that at most 6 offer such courses.
- f. Is it more likely that 0 or that 13 will offer such courses? Use numbers to justify your answer numerically and answer in a complete sentence.

**Exercise 4.9**

(*Solution on p. 82.*)

A school newspaper reporter decides to randomly survey 12 students to see if they will attend Tet festivities this year. Based on past years, she knows that 18% of students attend Tet festivities. We are interested in the number of students who will attend the festivities.

- d. How many of the 12 students do we expect to attend the festivities?
- e. Find the probability that at most 4 students will attend.
- f. Find the probability that more than 2 students will attend.

**Exercise 4.10**

Suppose that about 85% of graduating students attend their graduation. A group of 22 graduating students is randomly chosen.

- d. How many are expected to attend their graduation?
- e. Find the probability that 17 or 18 attend.
- f. Based on numerical values, would you be surprised if all 22 attended graduation? Justify your answer numerically.

**Exercise 4.11**

(*Solution on p. 82.*)

At The Fencing Center, 60% of the fencers use the foil as their main weapon. We randomly survey 25 fencers at The Fencing Center. We are interested in the numbers that do **not** use the foil as their main weapon.

- d. How many are expected to **not** use the foil as their main weapon?
- e. Find the probability that six do **not** use the foil as their main weapon.
- f. Based on numerical values, would you be surprised if all 25 did **not** use foil as their main weapon? Justify your answer numerically.

**Exercise 4.12**

Approximately 8% of students at a local high school participate in after-school sports all four years of high school. A group of 60 seniors is randomly chosen. Of interest is the number that participated in after-school sports all four years of high school.

- d. How many seniors are expected to have participated in after-school sports all four years of high school?

---

<sup>3</sup><http://nces.ed.gov/pubs2003/2003017.pdf>

- e. Based on numerical values, would you be surprised if none of the seniors participated in after-school sports all four years of high school? Justify your answer numerically.
- f. Based upon numerical values, is it more likely that 4 or that 5 of the seniors participated in after-school sports all four years of high school? Justify your answer numerically.

**Exercise 4.13**

(Solution on p. 82.)

The chance of having an extra fortune in a fortune cookie is about 3%. Given a bag of 144 fortune cookies, we are interested in the number of cookies with an extra fortune. Two distributions may be used to solve this problem. Use one distribution to solve the problem.

- d. How many cookies do we expect to have an extra fortune?
  - e. Find the probability that none of the cookies have an extra fortune.
  - f. Find the probability that more than 3 have an extra fortune.
  - g. As  $n$  increases, what happens involving the probabilities using the two distributions?
- Explain in complete sentences.

**Exercise 4.14**

There are two games played for Chinese New Year and Vietnamese New Year. They are almost identical. In the Chinese version, fair dice with numbers 1, 2, 3, 4, 5, and 6 are used, along with a board with those numbers. In the Vietnamese version, fair dice with pictures of a gourd, fish, rooster, crab, crayfish, and deer are used. The board has those six objects on it, also. We will play with bets being \$1. The player places a bet on a number or object. The “house” rolls three dice. If none of the dice show the number or object that was bet, the house keeps the \$1 bet. If one of the dice shows the number or object bet (and the other two do not show it), the player gets back his \$1 bet, plus \$1 profit. If two of the dice show the number or object bet (and the third die does not show it), the player gets back his \$1 bet, plus \$2 profit. If all three dice show the number or object bet, the player gets back his \$1 bet, plus \$3 profit.

Let  $X$  = number of matches and  $Y$  = profit per game.

- d. List the values that  $Y$  may take on. Then, construct one PDF table that includes both  $X$  &  $Y$  and their probabilities.
- e. Calculate the average expected matches over the long run of playing this game for the player.
- f. Calculate the average expected earnings over the long run of playing this game for the player.
- g. Determine who has the advantage, the player or the house.

**Exercise 4.15**

(Solution on p. 82.)

According to the South Carolina Department of Mental Health web site, for every 200 U.S. women, the average number who suffer from anorexia is one (<http://www.state.sc.us/dmh/anorexia/statistics.htm><sup>4</sup>). Out of a randomly chosen group of 600 U.S. women:

- d. How many are expected to suffer from anorexia?
- e. Find the probability that no one suffers from anorexia.
- f. Find the probability that more than four suffer from anorexia.

**Exercise 4.16**

The average number of children of middle-aged Japanese couples is 2.09 (Source: *The Yomiuri Shimbun*, June 28, 2006). Suppose that one middle-aged Japanese couple is randomly chosen.

---

<sup>4</sup><http://www.state.sc.us/dmh/anorexia/statistics.htm>

- d. Find the probability that they have no children.
- e. Find the probability that they have fewer children than the Japanese average.
- f. Find the probability that they have more children than the Japanese average .

**Exercise 4.17**

(Solution on p. 83.)

The average number of children per Spanish couples was 1.34 in 2005. Suppose that one Spanish couple is randomly chosen. (Source: [http://www.typicallyspanish.com/news/publish/article\\_4897.shtml](http://www.typicallyspanish.com/news/publish/article_4897.shtml)<sup>5</sup>, June 16, 2006).

- d. Find the probability that they have no children.
- e. Find the probability that they have fewer children than the Spanish average.
- f. Find the probability that they have more children than the Spanish average .

**Exercise 4.18**

Fertile (female) cats produce an average of 3 litters per year. (Source: *The Humane Society of the United States*). Suppose that one fertile, female cat is randomly chosen. In one year, find the probability she produces:

- d. No litters.
- e. At least 2 litters.
- f. Exactly 3 litters.

**Exercise 4.19**

(Solution on p. 83.)

A consumer looking to buy a used red Miata car will call dealerships until she finds a dealership that carries the car. She estimates the probability that any independent dealership will have the car will be 28%. We are interested in the number of dealerships she must call.

- d. On average, how many dealerships would we expect her to have to call until she finds one that has the car?
- e. Find the probability that she must call at most 4 dealerships.
- f. Find the probability that she must call 3 or 4 dealerships.

**Exercise 4.20**

Suppose that the probability that an adult in America will watch the Super Bowl is 40%. Each person is considered independent. We are interested in the number of adults in America we must survey until we find one who will watch the Super Bowl.

- d. How many adults in America do you expect to survey until you find one who will watch the Super Bowl?
- e. Find the probability that you must ask 7 people.
- f. Find the probability that you must ask 3 or 4 people.

**Exercise 4.21**

(Solution on p. 83.)

A group of Martial Arts students is planning on participating in an upcoming demonstration. 6 are students of Tae Kwon Do; 7 are students of Shotokan Karate. Suppose that 8 students are randomly picked to be in the first demonstration. We are interested in the number of Shotokan Karate students in that first demonstration.

- d. How many Shotokan Karate students do we expect to be in that first demonstration?
- e. Find the probability that 4 students of Shotokan Karate are picked.
- f. Find the probability that no more than 6 students of Shotokan Karate are picked.

---

<sup>5</sup>[http://www.typicallyspanish.com/news/publish/article\\_4897.shtml](http://www.typicallyspanish.com/news/publish/article_4897.shtml)

**Exercise 4.22**

The chance of a IRS audit for a tax return with over \$25,000 in income is about 2% per year. We are interested in the expected number of audits a person with that income has in a 20 year period. Assume each year is independent.

- d. How many audits are expected in a 20 year period?
- e. Find the probability that a person is not audited at all.
- f. Find the probability that a person is audited more than twice.

**Exercise 4.23**

*(Solution on p. 83.)*

Refer to the previous problem. Suppose that 100 people with tax returns over \$25,000 are randomly picked. We are interested in the number of people audited in 1 year. One way to solve this problem is by using the Binomial Distribution. Since  $n$  is large and  $p$  is small, another discrete distribution could be used to solve the following problems. Solve the following questions (d-f) using that distribution.

- d. How many are expected to be audited?
- e. Find the probability that no one was audited.
- f. Find the probability that more than 2 were audited.

**Exercise 4.24**

Suppose that a technology task force is being formed to study technology awareness among instructors. Assume that 10 people will be randomly chosen to be on the committee from a group of 28 volunteers, 20 who are technically proficient and 8 who are not. We are interested in the number on the committee who are **not** technically proficient.

- d. How many instructors do you expect on the committee who are **not** technically proficient?
- e. Find the probability that at least 5 on the committee are not technically proficient.
- f. Find the probability that at most 3 on the committee are not technically proficient.

**Exercise 4.25**

*(Solution on p. 83.)*

Refer back to Exercise 4.15.12. Solve this problem again, using a different, though still acceptable, distribution.

**Exercise 4.26**

Suppose that 9 Massachusetts athletes are scheduled to appear at a charity benefit. The 9 are randomly chosen from 8 volunteers from the Boston Celtics and 4 volunteers from the New England Patriots. We are interested in the number of Patriots picked.

- d. Is it more likely that there will be 2 Patriots or 3 Patriots picked?
- e. What is the probability that all of the volunteers will be from the Celtics?
- f. Is it more likely that more of the volunteers will be from the Patriots or from the Celtics?  
How do you know?

**Exercise 4.27**

*(Solution on p. 83.)*

On average, Pierre, an amateur chef, drops 3 pieces of egg shell into every 2 batters of cake he makes. Suppose that you buy one of his cakes.

- d. On average, how many pieces of egg shell do you expect to be in the cake?
- e. What is the probability that there will not be any pieces of egg shell in the cake?
- f. Let's say that you buy one of Pierre's cakes each week for 6 weeks. What is the probability that there will not be any egg shell in any of the cakes?
- g. Based upon the average given for Pierre, is it possible for there to be 7 pieces of shell in the cake? Why?

**Exercise 4.28**

It has been estimated that only about 30% of California residents have adequate earthquake supplies. Suppose we are interested in the number of California residents we must survey until we find a resident who does **not** have adequate earthquake supplies.

- d. What is the probability that we must survey just 1 or 2 residents until we find a California resident who does not have adequate earthquake supplies?
- e. What is the probability that we must survey at least 3 California residents until we find a California resident who does not have adequate earthquake supplies?
- f. How many California residents do you expect to need to survey until you find a California resident who **does not** have adequate earthquake supplies?
- g. How many California residents do you expect to need to survey until you find a California resident who **does** have adequate earthquake supplies?

**Exercise 4.29***(Solution on p. 83.)*

Refer to the above problem. Suppose you randomly survey 11 California residents. We are interested in the number who have adequate earthquake supplies.

- d. What is the probability that at least 8 have adequate earthquake supplies?
- e. Is it more likely that none or that all of the residents surveyed will have adequate earthquake supplies? Why?
- f. How many residents do you expect will have adequate earthquake supplies?

The next 3 questions refer to the following: In one of its Spring catalogs, L.L. Bean® advertised footwear on 29 of its 192 catalog pages.

**Exercise 4.30**

Suppose we randomly survey 20 pages. We are interested in the number of pages that advertise footwear. Each page may be picked at most once.

- d. How many pages do you expect to advertise footwear on them?
- e. Is it probable that all 20 will advertise footwear on them? Why or why not?
- f. What is the probability that less than 10 will advertise footwear on them?

**Exercise 4.31***(Solution on p. 84.)*

Suppose we randomly survey 20 pages. We are interested in the number of pages that advertise footwear. This time, each page may be picked more than once.

- d. How many pages do you expect to advertise footwear on them?
- e. Is it probable that all 20 will advertise footwear on them? Why or why not?
- f. What is the probability that less than 10 will advertise footwear on them?
- g. Suppose that a page may be picked more than once. We are interested in the number of pages that we must randomly survey until we find one that has footwear advertised on it. Define the random variable  $X$  and give its distribution.
- h. Do you expect to survey more than 10 pages in order to find one that advertises footwear on it? Why?
- i. What is the probability that you only need to survey at most 3 pages in order to find one that advertises footwear on it?
- j. How many pages do you expect to need to survey in order to find one that advertises footwear?

**Exercise 4.32**

Suppose that you roll a fair die until each face has appeared at least once. It does not matter in what order the numbers appear. Find the expected number of rolls you must make until each face has appeared at least once.

### 4.2.2 Try these multiple choice problems.

**For the next three problems:** The probability that the San Jose Sharks will win any given game is 0.3694 based on their 13 year win history of 382 wins out of 1034 games played (as of a certain date). Their 2005 schedule for November contains 12 games. Let  $X =$  number of games won in November 2005

#### Exercise 4.33

(*Solution on p. 84.*)

The expected number of wins for the month of November 2005 is:

- A. 1.67
- B. 12
- C.  $\frac{382}{1034}$
- D. 4.43

#### Exercise 4.34

(*Solution on p. 84.*)

What is the probability that the San Jose Sharks win 6 games in November?

- A. 0.1476
- B. 0.2336
- C. 0.7664
- D. 0.8903

#### Exercise 4.35

(*Solution on p. 84.*)

Find the probability that the San Jose Sharks win at least 5 games in November.

- A. 0.3694
- B. 0.5266
- C. 0.4734
- D. 0.2305

**For the next three questions:** The average number of times per week that Mrs. Plum's cats wake her up at night because they want to play is 10. We are interested in the number of times her cats wake her up each week.

#### Exercise 4.36

(*Solution on p. 84.*)

In words, the random variable  $X =$

- A. The number of times Mrs. Plum's cats wake her up each week
- B. The number of times Mrs. Plum's cats wake her up each hour
- C. The number of times Mrs. Plum's cats wake her up each night
- D. The number of times Mrs. Plum's cats wake her up

#### Exercise 4.37

(*Solution on p. 84.*)

Find the probability that her cats will wake me up no more than 5 times next week.

- A. 0.5000
- B. 0.9329
- C. 0.0378
- D. 0.0671

#### Exercise 4.38

(*Solution on p. 84.*)

People visiting video rental stores often rent more than one DVD at a time. The probability distribution for DVD rentals per customer at Video To Go is given below. There is 5 video limit per customer at this store, so nobody ever rents more than 5 DVDs.

X	0	1	2	3	4	5
P(X)	0.03	0.50	0.24	?	0.07	0.04

**Table 4.4**

- Describe the random variable X in words.
- Find the probability that a customer rents three DVDs.
- Find the probability that a customer rents at least 4 DVDs. Write your answer using proper notation.
- Find the probability that a customer rents at most 2 DVDs. Write your answer using proper notation.

Another shop, Entertainment Headquarters, rents DVDs and videogames. The probability distribution for DVD rentals per customer at this shop is given below. They also have a 5 DVD limit per customer.

X)	0	1	2	3	4	5
P(X)	0.35	0.25	0.20	0.10	0.05	0.05

**Table 4.5**

- At which store is the expected number of DVDs rented per customer higher?
- If Video to Go estimates that they will have 300 customers next week, how many DVDs do they expect to rent next week? Answer in sentence form.
- If Video to Go expects 300 customers next week and Entertainment HQ projects that they will have 420 customers, for which store is the expected number of DVD rentals for next week higher? Explain.
- Which of the two video stores experiences more variation in the number of DVD rentals per customer? How do you know that?

**Exercise 4.39***(Solution on p. 84.)*

A game involves selecting a card from a deck of cards and tossing a coin. The deck has 52 cards and 12 cards are "face cards" (Jack, Queen, or King) The coin is a fair coin and is equally likely to land on Heads or Tails

- If the card is a face card and the coin lands on Heads, you win \$6
- If the card is a face card and the coin lands on Tails, you win \$2
- If the card is not a face card, you lose \$2, no matter what the coin shows.

- Find the expected value for this game (expected net gain or loss).
- Explain what your calculations indicate about your long-term average profits and losses on this game.
- Should you play this game to win money?

**Exercise 4.40***(Solution on p. 84.)*

You buy a lottery ticket to a lottery that costs \$10 per ticket. There are only 100 tickets available be sold in this lottery. In this lottery there is one \$500 prize, 2 \$100 prizes and 4 \$25 prizes. Find your expected gain or loss.

**Exercise 4.41**

(Solution on p. 85.)

A student takes a 10 question true-false quiz, but did not study and randomly guesses each answer. Find the probability that the student passes the quiz with a grade of **at least** 70% of the questions correct.

**Exercise 4.42**

(Solution on p. 85.)

A student takes a 32 question multiple choice exam, but did not study and randomly guesses each answer. Each question has 3 possible choices for the answer. Find the probability that the student guesses **more than** 75% of the questions correctly.

**Exercise 4.43**

(Solution on p. 85.)

Suppose that you are performing the probability experiment of rolling one die. Let F be the event of rolling a "4" or a "5". You are interested in how many times you need to roll the die in order to obtain the first "4 or 5" as the outcome.

- p = probability of success (event F occurs)
  - q = probability of failure (event F does not occur)
- a. Write the description of the random variable X. What are the values that X can take on? Find the values of p and q. What is the appropriate probability distribution for X?
  - b. Find the probability that the first occurrence of event F ("4" or "5") is on the first or second trial.
  - c. Find the probability that more than 4 trials are needed to obtain the first "4" or "5" when rolling the die.

### 4.3 Review Questions<sup>6</sup>

The next two questions refer to the following:

A recent poll concerning credit cards found that 35 percent of respondents use a credit card that gives them a mile of air travel for every dollar they charge. Thirty percent of the respondents charge more than \$2000 per month. Of those respondents who charge more than \$2000, 80 percent use a credit card that gives them a mile of air travel for every dollar they charge.

**Exercise 4.44: REVIEW QUESTION 1**

(Solution on p. 85.)

What is the probability that a randomly selected respondent expected to spend more than \$2000 AND use a credit card that gives them a mile of air travel for every dollar they charge?

- A. (0.30) (0.35)
- B. (0.80) (0.35)
- C. (0.80) (0.30)
- D. (0.80)

**Exercise 4.45: REVIEW QUESTION 2**

(Solution on p. 85.)

Based upon the above information, are using a credit card that gives a mile of air travel for each dollar spent AND charging more than \$2000 per month independent events?

- A. Yes
- B. No, and they are not mutually exclusive either
- C. No, but they are mutually exclusive

---

<sup>6</sup>This content is available online at <<http://cnx.org/content/m19021/1.1/>>.

- D. Not enough information given to determine the answer

**Exercise 4.46: REVIEW QUESTION 3**

(*Solution on p. 85.*)

A sociologist wants to know the opinions of employed adult women about government funding for day care. She obtains a list of 520 members of a local business and professional women's club and mails a questionnaire to 100 of these women selected at random. 68 questionnaires are returned. What is the population in this study?

- A. All employed adult women
- B. All the members of a local business and professional women's club
- C. The 100 women who received the questionnaire
- D. All employed women with children

The next two questions refer to the following: An article from The San Jose Mercury News was concerned with the racial mix of the 1500 students at Prospect High School in Saratoga, CA. The table summarizes the results. (Male and female values are approximate.)

			Ethnic Group		
Gender	White	Asian	Hispanic	Black	American Indian
Male	400	168	115	35	16
Female	440	132	140	40	14

Table 4.6

**Exercise 4.47: REVIEW QUESTION 4**

(*Solution on p. 85.*)

Find the probability that a student is Asian or Male.

**Exercise 4.48: REVIEW QUESTION 5**

(*Solution on p. 85.*)

Find the probability that a student is Black given that the student is Female.

**Exercise 4.49: REVIEW QUESTION 6**

(*Solution on p. 85.*)

A sample of pounds lost, in a certain month, by individual members of a weight reducing clinic produced the following statistics:

- Mean = 5 lbs.
- Median = 4.5 lbs.
- Mode = 4 lbs.
- Standard deviation = 3.8 lbs.
- First quartile = 2 lbs.
- Third quartile = 8.5 lbs.

The correct statement is:

- A. One fourth of the members lost exactly 2 pounds.
- B. The middle fifty percent of the members lost from 2 to 8.5 lbs.
- C. Most people lost 3.5 to 4.5 lbs.
- D. All of the choices above are correct.

**Exercise 4.50: REVIEW QUESTION 7**

(*Solution on p. 85.*)

What does it mean when a data set has a standard deviation equal to zero?

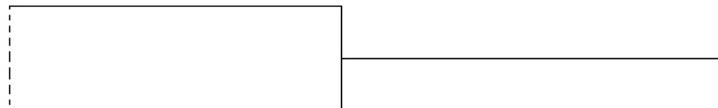
- A. All values of the data appear with the same frequency.
- B. The mean of the data is also zero.

- C. All of the data have the same value.
- D. There are no data to begin with.

**Exercise 4.51: REVIEW QUESTION 8**

(Solution on p. 85.)

The statement that best describes the illustration below is:

**Figure 4.1**

- A. The mean is equal to the median.
- B. There is no first quartile.
- C. The lowest data value is the median.
- D. The median equals  $\frac{(Q_1+Q_3)}{2}$

**Exercise 4.52: REVIEW QUESTION 9**

(Solution on p. 85.)

According to a recent article (San Jose Mercury News) the average number of babies born with significant hearing loss (deafness) is approximately 2 per 1000 babies in a healthy baby nursery. The number climbs to an average of 30 per 1000 babies in an intensive care nursery.

Suppose that 1000 babies from healthy nursery babies were surveyed. Find the probability that exactly 2 babies were born deaf.

**Exercise 4.53: REVIEW QUESTION 10**

(Solution on p. 86.)

A “friend” offers you the following “deal.” For a \$10 fee, you may pick an envelope from a box containing 100 seemingly identical envelopes. However, each envelope contains a coupon for a free gift.

- 10 of the coupons are for a free gift worth \$6.
- 80 of the coupons are for a free gift worth \$8.
- 6 of the coupons are for a free gift worth \$12.
- 4 of the coupons are for a free gift worth \$40.

Based upon the financial gain or loss over the long run, should you play the game?

- A. Yes, I expect to come out ahead in money.
- B. No, I expect to come out behind in money.
- C. It doesn’t matter. I expect to break even.

The next four questions refer to the following: Recently, a nurse commented that when a patient calls the medical advice line claiming to have **the flu**, the chance that he/she truly has **the flu** (and not just a nasty cold) is only about 4%. Of the next 25 patients calling in claiming to have **the flu**, we are interested in how many actually have **the flu**.

**Exercise 4.54: REVIEW QUESTION 11**

(Solution on p. 86.)

Define the Random Variable and list its possible values.

**Exercise 4.55: REVIEW QUESTION 12***(Solution on p. 86.)*

State the distribution of  $X$ .

**Exercise 4.56: REVIEW QUESTION 13***(Solution on p. 86.)*

Find the probability that at least 4 of the 25 patients actually have **the flu**.

**Exercise 4.57: REVIEW QUESTION 14***(Solution on p. 86.)*

On average, for every 25 patients calling in, how many do you expect to have **the flu**?

The next two questions refer to the following: Different types of writing can sometimes be distinguished by the number of letters in the words used. A student interested in this fact wants to study the number of letters of words used by Tom Clancy in his novels. She opens a Clancy novel at random and records the number of letters of the first 250 words on the page.

**Exercise 4.58: REVIEW QUESTION 15***(Solution on p. 86.)*

What kind of data was collected?

- A. qualitative
- B. quantitative - continuous
- C. quantitative – discrete

**Exercise 4.59: REVIEW QUESTION 16***(Solution on p. 86.)*

What is the population under study?

## Solutions to Exercises in Chapter 4

### Solution to Exercise 4.1 (p. 68)

- a. 0.1
- b. 1.6

### Solution to Exercise 4.3 (p. 69)

- b. \$200,000;\$600,000;\$400,000
- c. third investment
- d. first investment
- e. second investment

### Solution to Exercise 4.5 (p. 69)

- a. 0.2
- c. 2.35
- d. 2-3 children

### Solution to Exercise 4.7 (p. 70)

- a.  $X =$  the number of dice that show a 1
- b. 0,1,2,3,4,5,6
- c.  $X \sim B\left(6, \frac{1}{6}\right)$
- d. 1
- e. 0.00002
- f. 3 dice

### Solution to Exercise 4.9 (p. 71)

- a.  $X =$  the number of students that will attend Tet.
- b. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
- c.  $X \sim B(12, 0.18)$
- d. 2.16
- e. 0.9511
- f. 0.3702

### Solution to Exercise 4.11 (p. 71)

- a.  $X =$  the number of fencers that do **not** use foil as their main weapon
- b. 0, 1, 2, 3,... 25
- c.  $X \sim B(25, 0.40)$
- d. 10
- e. 0.0442
- f. Yes

### Solution to Exercise 4.13 (p. 72)

- a.  $X =$  the number of fortune cookies that have an extra fortune
- b. 0, 1, 2, 3,... 144
- c.  $X \sim B(25, 0.40)$  or  $P(4.32)$
- d. 4.32
- e. 0.0124 or 0.0133
- f. 0.6300 or 0.6264

### Solution to Exercise 4.15 (p. 72)

- a.  $X$  = the number of women that suffer from anorexia
- b. 0, 1, 2, 3,... 600 (can leave off 600)
- c.  $X \sim P(3)$
- d. 3
- e. 0.0498
- f. 0.1847

**Solution to Exercise 4.17 (p. 73)**

- a.  $X$  = the number of children for a Spanish couple
- b. 0, 1, 2, 3,...
- c.  $X \sim P(1.34)$
- d. 0.2618
- e. 0.6217
- f. 0.3873

**Solution to Exercise 4.19 (p. 73)**

- a.  $X$  = the number of dealers she calls until she finds one with a used red Miata
- b. 0, 1, 2, 3,...
- c.  $X \sim G(0.28)$
- d. 3.57
- e. 0.7313
- f. 0.2497

**Solution to Exercise 4.21 (p. 73)**

- d. 4.31
- e. 0.4079
- f. 0.9953

**Solution to Exercise 4.23 (p. 74)**

- d. 2
- e. 0.1353
- f. 0.3233

**Solution to Exercise 4.25 (p. 74)**

- a.  $X$  = the number of seniors that participated in after-school sports all 4 years of high school
- b. 0, 1, 2, 3,... 60
- c.  $X \sim P(4.8)$
- d. 4.8
- e. Yes
- f. 4

**Solution to Exercise 4.27 (p. 74)**

- a.  $X$  = the number of shell pieces in one cake
- b. 0, 1, 2, 3,...
- c.  $X \sim P(1.5)$
- d. 1.5
- e. 0.2231
- f. 0.0001
- g. Yes

**Solution to Exercise 4.29 (p. 75)**

- d. 0.0043
- e. none
- f. 3.3

**Solution to Exercise 4.31 (p. 75)**

- d. 3.02
- e. No
- f. 0.9997
- h. 0.2291
- i. 0.3881
- j. 6.6207 pages

**Solution to Exercise 4.33 (p. 76)**

D: 4.43

**Solution to Exercise 4.34 (p. 76)**

A: 0.1476

**Solution to Exercise 4.35 (p. 76)**

C: 0.4734

**Solution to Exercise 4.36 (p. 76)**

A: The number of times Mrs. Plum's cats wake her up each week

**Solution to Exercise 4.37 (p. 76)**

D: 0.0671

**Solution to Exercise 4.38 (p. 76)**

Solution will be posted on the instructor's website for this class.

**Solution to Exercise 4.39 (p. 77)**The variable of interest is  $X$  = net gain or loss, in dollars

The face cards J, Q, K (Jack, Queen, King). There are  $(3)(4) = 12$  face cards and  $52 - 12 = 40$  cards that are not face cards.

We first need to construct the probability distribution for  $X$ . We use the card and coin events to determine the probability for each outcome, but we use the monetary value of  $X$  to determine the expected value.

Card Event	$\$X$ net gain or loss	$P(X)$
Face Card and Heads	6	$(12/52)(1/2) = 6/52$
Face Card and Tails	2	$(12/52)(1/2) = 6/52$
(Not Face Card) and (H or T)	-2	$(40/52)(1) = 40/52$

**Table 4.7**

- Expected value =  $(6)(6/52) + (2)(6/52) + (-2)(40/52) = -32/52$
- Expected value =  $-\$0.62$ , rounded to the nearest cent
- If you play this game repeatedly, over a long number of games, you would expect to lose 62 cents per game, on average.
- You should not play this game to win money because the expected value indicates an expected average loss.

**Solution to Exercise 4.40 (p. 77)**Start by writing the probability distribution.  $X$  is net gain or loss = prize (if any) less \$10 cost of ticket

$X = \$ \text{ net gain or loss}$	$P(X)$
$\$500 - \$10 = \$490$	$1/100$
$\$100 - \$10 = \$90$	$2/100$
$\$25 - \$10 = \$15$	$4/100$
$\$0 - \$10 = -\$10$	$93/100$

**Table 4.8**

Expected Value =  $(490)(1/100) + (90)(2/100) + (15)(4/100) + (-10)(93/100) = -\$2$ . There is an expected loss of \$2 per ticket, on average.

#### Solution to Exercise 4.41 (p. 78)

- $X = \text{number of questions answered correctly}$
- $X \sim B(10, 0.5)$
- We are interested in AT LEAST 70% of 10 questions correct. 70% of 10 is 7. We want to find the probability that  $X$  is greater than or equal to 7. The event "at least 7" is the complement of "less than or equal to 6".
- Using your calculator's distribution menu:  $1 - \text{binomcdf}(10, .5, 6)$  gives 0.171875
- The probability of getting at least 70% of the 10 questions correct when randomly guessing is approximately 0.172

#### Solution to Exercise 4.42 (p. 78)

- $X = \text{number of questions answered correctly}$
- $X \sim B(32, 1/3)$
- We are interested in MORE THAN 75% of 32 questions correct. 75% of 32 is 24. We want to find  $P(X > 24)$ . The event "more than 24" is the complement of "less than or equal to 24".
- Using your calculator's distribution menu:  $1 - \text{binomcdf}(32, 1/3, 24)$
- $P(X > 24) = 0.00000026761$
- The probability of getting more than 75% of the 32 questions correct when randomly guessing is very small and practically zero.

#### Solution to Exercise 4.43 (p. 78)

Solution will be posted on the instructor's website for this class.

#### Solution to Exercise 4.44 (p. 78)

REVIEW QUESTION 1 Solution : C

#### Solution to Exercise 4.45 (p. 78)

REVIEW QUESTION 2 Solution : B

#### Solution to Exercise 4.46 (p. 79)

REVIEW QUESTION 3 Solution : A

#### Solution to Exercise 4.47 (p. 79)

REVIEW QUESTION 4 Solution : 0.5773

#### Solution to Exercise 4.48 (p. 79)

REVIEW QUESTION 5 Solution : 0.0522

#### Solution to Exercise 4.49 (p. 79)

REVIEW QUESTION 6 Solution : B

#### Solution to Exercise 4.50 (p. 79)

REVIEW QUESTION 7 Solution : C

#### Solution to Exercise 4.51 (p. 80)

REVIEW QUESTION 8 Solution : C

**Solution to Exercise 4.52 (p. 80)**

REVIEW QUESTION 9 Solution : 0.2709

**Solution to Exercise 4.53 (p. 80)**

REVIEW QUESTION 10 Solution : B

**Solution to Exercise 4.54 (p. 80)****REVIEW QUESTION 11 Solution**

$X$  = the number of patients calling in claiming to have **the flu**, who actually have **the flu**.  $X = 0, 1, 2, \dots, 25$

**Solution to Exercise 4.55 (p. 81)**

REVIEW QUESTION 12 Solution : B (25, 0.04)

**Solution to Exercise 4.56 (p. 81)**

REVIEW QUESTION 13 Solution : 0.0165

**Solution to Exercise 4.57 (p. 81)**

REVIEW QUESTION 14 Solution : 1 patient

**Solution to Exercise 4.58 (p. 81)**

REVIEW QUESTION 15 Solution : C

**Solution to Exercise 4.59 (p. 81)**

REVIEW QUESTION 16 Solution : All words used by Tom Clancy in his novels

$$\text{Percentile, } k: k = \frac{\ln(1-\text{AreaToTheLeft})}{-\mu}$$

## 5.2 Homework<sup>2</sup>

For each probability and percentile problem, DRAW THE PICTURE!

### Exercise 5.1

Consider the following experiment. You are one of 100 people enlisted to take part in a study to determine the percent of nurses in America with an R.N. (registered nurse) degree. You ask nurses if they have an R.N. degree. The nurses answer “yes” or “no.” You then calculate the percentage of nurses with an R.N. degree. You give that percentage to your supervisor.

- a. What part of the experiment will yield discrete data?
- b. What part of the experiment will yield continuous data?

### Exercise 5.2

When age is rounded to the nearest year, do the data stay continuous, or do they become discrete? Why?

### Exercise 5.3

(Solution on p. 96.)

Births are approximately uniformly distributed between the 52 weeks of the year. They can be said to follow a Uniform Distribution from 1 – 53 (spread of 52 weeks).

- a.  $X \sim$
- b. Graph the probability distribution.
- c.  $f(x) =$
- d.  $\mu =$
- e.  $\sigma =$
- f. Find the probability that a person is born at the exact moment week 19 starts. That is, find  $P(X = 19) =$
- g.  $P(2 < X < 31) =$
- h. Find the probability that a person is born after week 40.
- i.  $P(12 < X | X < 28) =$
- j. Find the 70th percentile.
- k. Find the minimum for the upper quarter.

### Exercise 5.4

A random number generator picks a number from 1 to 9 in a uniform manner.

- a.  $X \sim$
- b. Graph the probability distribution.
- c.  $f(x) =$
- d.  $\mu =$
- e.  $\sigma =$
- f.  $P(3.5 < X < 7.25) =$
- g.  $P(X > 5.67) =$
- h.  $P(X > 5 | X > 3) =$
- i. Find the 90th percentile.

### Exercise 5.5

(Solution on p. 96.)

The time (in minutes) until the next bus departs a major bus depot follows a distribution with  $f(x) = \frac{1}{20}$  where  $x$  goes from 25 to 45 minutes.

---

<sup>2</sup>This content is available online at <<http://cnx.org/content/m16807/1.12/>>.

- a.  $X =$
- b.  $X \sim$
- c. Graph the probability distribution.
- d. The distribution is \_\_\_\_\_ (name of distribution). It is \_\_\_\_\_ (discrete or continuous).
- e.  $\mu =$
- f.  $\sigma =$
- g. Find the probability that the time is at most 30 minutes. Sketch and label a graph of the distribution. Shade the area of interest. Write the answer in a probability statement.
- h. Find the probability that the time is between 30 and 40 minutes. Sketch and label a graph of the distribution. Shade the area of interest. Write the answer in a probability statement.
- i.  $P(25 < X < 55) =$  \_\_\_\_\_. State this in a probability statement (similar to g and h), draw the picture, and find the probability.
- j. Find the 90th percentile. This means that 90% of the time, the time is less than \_\_\_\_ minutes.
- k. Find the 75th percentile. In a complete sentence, state what this means. (See j.)
- l. Find the probability that the time is more than 40 minutes given (or knowing that) it is at least 30 minutes.

### Exercise 5.6

According to a study by Dr. John McDougall of his live-in weight loss program at St. Helena Hospital, the people who follow his program lose between 6 and 15 pounds a month until they approach trim body weight. Let's suppose that the weight loss is uniformly distributed. We are interested in the weight loss of a randomly selected individual following the program for one month. (Source: **The McDougall Program for Maximum Weight Loss** by John A. McDougall, M.D.)

- a.  $X =$
- b.  $X \sim$
- c. Graph the probability distribution.
- d.  $f(x) =$
- e.  $\mu =$
- f.  $\sigma =$
- g. Find the probability that the individual lost more than 10 pounds in a month.
- h. Suppose it is known that the individual lost more than 10 pounds in a month. Find the probability that he lost less than 12 pounds in the month.
- i.  $P(7 < X < 13 | X > 9) =$  \_\_\_\_\_. State this in a probability question (similar to g and h), draw the picture, and find the probability.

### Exercise 5.7

*(Solution on p. 96.)*

A subway train on the Red Line arrives every 8 minutes during rush hour. We are interested in the length of time a commuter must wait for a train to arrive. The time follows a uniform distribution.

- a.  $X =$
- b.  $X \sim$
- c. Graph the probability distribution.
- d.  $f(x) =$
- e.  $\mu =$
- f.  $\sigma =$
- g. Find the probability that the commuter waits less than one minute.
- h. Find the probability that the commuter waits between three and four minutes.
- i. 60% of commuters wait more than how long for the train? State this in a probability question (similar to g and h), draw the picture, and find the probability.

**Exercise 5.8**

The age of a first grader on September 1 at Garden Elementary School is uniformly distributed from 5.8 to 6.8 years. We randomly select one first grader from the class.

- a.  $X =$
- b.  $X \sim$
- c. Graph the probability distribution.
- d.  $f(x) =$
- e.  $\mu =$
- f.  $\sigma =$
- g. Find the probability that she is over 6.5 years.
- h. Find the probability that she is between 4 and 6 years.
- i. Find the 70th percentile for the age of first graders on September 1 at Garden Elementary School.

**Exercise 5.9**

(*Solution on p. 96.*)

Let  $X \sim \text{Exp}(0.1)$

- a. decay rate=
- b.  $\mu =$
- c. Graph the probability distribution function.
- d. On the above graph, shade the area corresponding to  $P(X < 6)$  and find the probability.
- e. Sketch a new graph, shade the area corresponding to  $P(3 < X < 6)$  and find the probability.
- f. Sketch a new graph, shade the area corresponding to  $P(X > 7)$  and find the probability.
- g. Sketch a new graph, shade the area corresponding to the 40th percentile and find the value.
- h. Find the average value of  $X$ .

**Exercise 5.10**

Suppose that the length of long distance phone calls, measured in minutes, is known to have an exponential distribution with the average length of a call equal to 8 minutes.

- a.  $X =$
- b. Is  $X$  continuous or discrete?
- c.  $X \sim$
- d.  $\mu =$
- e.  $\sigma =$
- f. Draw a graph of the probability distribution. Label the axes.
- g. Find the probability that a phone call lasts less than 9 minutes.
- h. Find the probability that a phone call lasts more than 9 minutes.
- i. Find the probability that a phone call lasts between 7 and 9 minutes.
- j. If 25 phone calls are made one after another, on average, what would you expect the total to be? Why?

**Exercise 5.11**

(*Solution on p. 96.*)

Suppose that the useful life of a particular car battery, measured in months, decays with parameter 0.025. We are interested in the life of the battery.

- a.  $X =$
- b. Is  $X$  continuous or discrete?
- c.  $X \sim$
- d. On average, how long would you expect 1 car battery to last?
- e. On average, how long would you expect 9 car batteries to last, if they are used one after another?
- f. Find the probability that a car battery lasts more than 36 months.

## Chapter 5

# Continuous Random Variables

### 5.1 Summary of the Uniform and Exponential Probability Distributions<sup>1</sup>

**Formula 5.1:** Uniform

$X$  = a real number between  $a$  and  $b$  (in some instances,  $X$  can take on the values  $a$  and  $b$ ).  $a$  = smallest  $X$ ;  $b$  = largest  $X$

$$X \sim U(a,b)$$

$$\text{The mean is } \mu = \frac{a+b}{2}$$

$$\text{The standard deviation is } \sigma = \sqrt{\frac{(b-a)^2}{12}}$$

$$\text{Probability density function: } f(X) = \frac{1}{b-a} \text{ for } a \leq X \leq b$$

$$\text{Area to the Left of } x: P(X < x) = (\text{base})(\text{height})$$

$$\text{Area to the Right of } x: P(X > x) = (\text{base})(\text{height})$$

$$\text{Area Between } c \text{ and } d: P(c < X < d) = (\text{base})(\text{height}) = (d - c)(\text{height}).$$

**Formula 5.2:** Exponential

$$X \sim Exp(m)$$

$X$  = a real number, 0 or larger.  $m$  = the parameter that controls the rate of decay or decline

The mean and standard deviation **are the same**.

$$\mu = \sigma = \frac{1}{m} \text{ and } m = \frac{1}{\mu} = \frac{1}{\sigma}$$

$$\text{The probability density function: } f(X) = m \cdot e^{-m \cdot X}, X \geq 0$$

$$\text{Area to the Left of } x: P(X < x) = 1 - e^{-m \cdot x}$$

$$\text{Area to the Right of } x: P(X > x) = e^{-m \cdot x}$$

$$\text{Area Between } c \text{ and } d: P(c < X < d) = P(X < d) - P(X < c) = \left(1 - e^{-m \cdot d}\right) - \left(1 - e^{-m \cdot c}\right) = e^{-m \cdot c} - e^{-m \cdot d}$$

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m16813/1.10/>>.

- g. 70% of the batteries last at least how long?

**Exercise 5.12**

The percent of persons (ages 5 and older) in each state who speak a language at home other than English is approximately exponentially distributed with a mean of 9.848 . Suppose we randomly pick a state. (Source: Bureau of the Census, U.S. Dept. of Commerce)

- a.  $X =$
- b. Is  $X$  continuous or discrete?
- c.  $X \sim$
- d.  $\mu =$
- e.  $\sigma =$
- f. Draw a graph of the probability distribution. Label the axes.
- g. Find the probability that the percent is less than 12.
- h. Find the probability that the percent is between 8 and 14.
- i. The percent of all individuals living in the United States who speak a language at home other than English is 13.8 .
  - i. Why is this number different from 9.848%?
  - ii. What would make this number higher than 9.848%?

**Exercise 5.13**

(*Solution on p. 97.*)

The time (in years) **after** reaching age 60 that it takes an individual to retire is approximately exponentially distributed with a mean of about 5 years. Suppose we randomly pick one retired individual. We are interested in the time after age 60 to retirement.

- a.  $X =$
- b. Is  $X$  continuous or discrete?
- c.  $X \sim$
- d.  $\mu =$
- e.  $\sigma =$
- f. Draw a graph of the probability distribution. Label the axes.
- g. Find the probability that the person retired after age 70.
- h. Do more people retire before age 65 or after age 65?
- i. In a room of 1000 people over age 80, how many do you expect will NOT have retired yet?

**Exercise 5.14**

The cost of all maintenance for a car during its first year is approximately exponentially distributed with a mean of \$150.

- a.  $X =$
- b.  $X \sim$
- c.  $\mu =$
- d.  $\sigma =$
- e. Draw a graph of the probability distribution. Label the axes.
- f. Find the probability that a car required over \$300 for maintenance during its first year.

### 5.2.1 Try these multiple choice problems

**The next three questions refer to the following information.** The average lifetime of a certain new cell phone is 3 years. The manufacturer will replace any cell phone failing within 2 years of the date of purchase. The lifetime of these cell phones is known to follow an exponential distribution.

**Exercise 5.15**

The decay rate is

- A. 0.3333
- B. 0.5000
- C. 2.0000
- D. 3.0000

(*Solution on p. 97.*)

**Exercise 5.16**

(*Solution on p. 97.*)

What is the probability that a phone will fail within 2 years of the date of purchase?

- A. 0.8647
- B. 0.4866
- C. 0.2212
- d. 0.9997

**Exercise 5.17**

(*Solution on p. 97.*)

What is the median lifetime of these phones (in years)?

- A. 0.1941
- B. 1.3863
- C. 2.0794
- D. 5.5452

**The next three questions refer to the following information.** The Sky Train from the terminal to the rental car and long term parking center is supposed to arrive every 8 minutes. The waiting times for the train are known to follow a uniform distribution.

**Exercise 5.18**

(*Solution on p. 97.*)

What is the average waiting time (in minutes)?

- A. 0.0000
- B. 2.0000
- C. 3.0000
- D. 4.0000

**Exercise 5.19**

(*Solution on p. 97.*)

Find the 30th percentile for the waiting times (in minutes).

- A. 2.0000
- B. 2.4000
- C. 2.750
- D. 3.000

**Exercise 5.20**

(*Solution on p. 97.*)

The probability of waiting more than 7 minutes given a person has waited more than 4 minutes is?

- A. 0.1250

- B. 0.2500
- C. 0.5000
- D. 0.7500

## 5.3 Review Questions<sup>3</sup>

**Review Questions 1 through 5 refer to the following study:** A recent study of mothers of junior high school children in Santa Clara County reported that 76% of the mothers are employed in paid positions. Of those mothers who are employed, 64% work full-time (over 35 hours per week), and 36% work part-time. However, out of all of the mothers in the population, 49% work full-time. The population under study is made up of mothers of junior high school children in Santa Clara County.

Let  $E$  =employed, Let  $F$  =full-time employment

**Exercise 5.21: REVIEW QUESTION 1**

(*Solution on p. 97.*)

- a. Find the percent of all mothers in the population that NOT employed.
- b. Find the percent of mothers in the population that are employed part-time.

**Exercise 5.22: REVIEW QUESTION 2**

(*Solution on p. 97.*)

The type of employment is considered to be what type of data?

**Exercise 5.23: REVIEW QUESTION 3**

(*Solution on p. 97.*)

In symbols, what does the 36% represent?

**Exercise 5.24: REVIEW QUESTION 4**

(*Solution on p. 97.*)

Find the probability that a randomly selected person from the population will be employed OR work full-time.

**Exercise 5.25: REVIEW QUESTION 5**

(*Solution on p. 97.*)

Based upon the above information, are being employed AND working part-time:

- a. mutually exclusive events? Why or why not?
- b. independent events? Why or why not?

**Review questions 6 and 7 refer to the following:** We randomly pick 10 mothers from the above population. We are interested in the number of the mothers that are employed. Let  $X$  =number of mothers that are employed.

**Exercise 5.26: REVIEW QUESTION 6**

(*Solution on p. 97.*)

State the distribution for  $X$ .

**Exercise 5.27: REVIEW QUESTION 7**

(*Solution on p. 97.*)

Find the probability that at least 6 are employed.

**Exercise 5.28: REVIEW QUESTION 8**

(*Solution on p. 97.*)

We expect the Statistics Discussion Board to have, on average, 14 questions posted to it per week. We are interested in the number of questions posted to it per day.

- a. Define  $X$ .
- b. What are the values that the random variable may take on?
- c. State the distribution for  $X$ .
- d. Find the probability that from 10 to 14 (inclusive) questions are posted to the Listserv on a randomly picked day.

---

<sup>3</sup>This content is available online at <<http://cnx.org/content/m19020/1.1/>>.

**Exercise 5.29: REVIEW QUESTION 9**

(Solution on p. 98.)

A person invests \$1000 in stock of a company that hopes to go public in 1 year.

- The probability that the person will lose all his money after 1 year (i.e. his stock will be worthless) is 35%.
- The probability that the person's stock will still have a value of \$1000 after 1 year (i.e. no profit and no loss) is 60%.
- The probability that the person's stock will increase in value by \$10,000 after 1 year (i.e. will be worth \$11,000) is 5%.

Find the expected PROFIT after 1 year.

**Exercise 5.30: REVIEW QUESTION 10**

(Solution on p. 98.)

Rachel's piano cost \$3000. The average cost for a piano is \$4000 with a standard deviation of \$2500. Becca's guitar cost \$550. The average cost for a guitar is \$500 with a standard deviation of \$200. Matt's drums cost \$600. The average cost for drums is \$700 with a standard deviation of \$100. Whose cost was lowest when compared to his or her own instrument? Justify your answer.

**Exercise 5.31: REVIEW QUESTION 11**

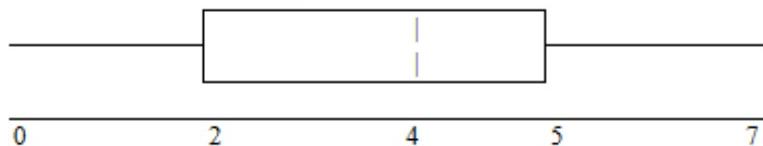
(Solution on p. 98.)

For the following data, which of the measures of central tendency would be the LEAST useful: mean, median, mode? Explain why. Which would be the MOST useful? Explain why.

4, 6, 6, 12, 18, 18, 18, 200

**Exercise 5.32: REVIEW QUESTION 12**

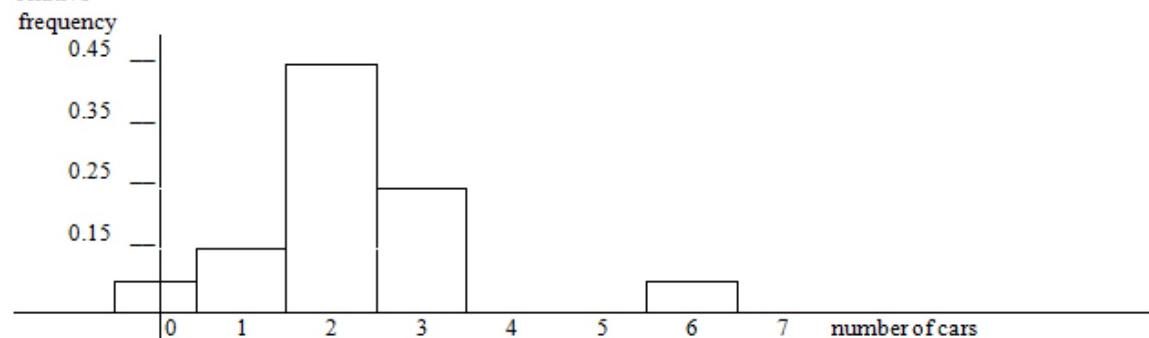
(Solution on p. 98.)



For each statement below, explain why each is either true or false.

- 25% of the data are at most 5.
- There is the same amount of data from 4 – 5 as there is from 5 – 7.
- There are no data values of 3.
- 50% of the data are 4.

**Review Questions 13 and 14 refer to the following:** 64 faculty members were asked the number of cars they owned (including spouse and children's cars). The results are given in the following graph:

**Exercise 5.33: REVIEW QUESTION 13**

(Solution on p. 98.)

Find the approximate number of responses that were "3."

**Exercise 5.34: REVIEW QUESTION 14**

(Solution on p. 98.)

Find the first, second and third quartiles. Use them to construct a box plot of the data.

Review Questions 15 and 16 refer to the following study done of the Girls soccer team "Snow Leopards":

Hair Style		Hair Color	
	blond	brown	black
ponytail	3	2	5
plain	2	2	1

Table 5.1

Suppose that one girl from the Snow Leopards is randomly selected.

**Exercise 5.35: REVIEW QUESTION 15**

(*Solution on p. 98.*)

Find the probability that the girl has black hair GIVEN that she wears a ponytail.

**Exercise 5.36: REVIEW QUESTION 16**

(*Solution on p. 98.*)

Find the probability that the girl wears her hair plain OR has brown hair.

**Exercise 5.37: REVIEW QUESTION 17**

(*Solution on p. 98.*)

Find the probability that the girl has blond hair AND that she wears her hair plain.

## Solutions to Exercises in Chapter 5

### Solution to Exercise 5.3 (p. 88)

- a.  $X \sim U(1, 53)$
- c.  $f(x) = \frac{1}{52}$  where  $1 \leq x \leq 53$
- d. 27
- e. 15.01
- f. 0
- g.  $\frac{29}{52}$
- h.  $\frac{13}{52}$
- i.  $\frac{16}{27}$
- j. 37.4
- k. 40

### Solution to Exercise 5.5 (p. 88)

- b.  $X \sim U(25, 45)$
- d. uniform; continuous
- e. 35 minutes
- f. 5.8 minutes
- g. 0.25
- h. 0.5
- i. 1
- j. 43 minutes
- k. 30 minutes
- l. 0.3333

### Solution to Exercise 5.7 (p. 89)

- b.  $X \sim U(0, 8)$
- d.  $f(x) = \frac{1}{8}$  where  $0 \leq X \leq 8$
- e. 4
- f. 2.31
- g.  $\frac{1}{8}$
- h.  $\frac{1}{8}$
- i. 3.2

### Solution to Exercise 5.9 (p. 90)

- a. 0.1
- b. 10
- d. 0.4512
- e. 0.1920
- f. 0.4966
- g. 5.11
- h. 10

### Solution to Exercise 5.11 (p. 90)

- c.  $X \sim \text{Exp}(0.025)$
- d. 40 months
- e. 360 months
- f. 0.4066
- g. 14.27

**Solution to Exercise 5.13 (p. 91)**

- c.  $X \sim \text{Exp} \left( \frac{1}{5} \right)$
- d. 5
- e. 5
- g. 0.1353
- h. Before
- i. 18.3

**Solution to Exercise 5.15 (p. 92)**

A

**Solution to Exercise 5.16 (p. 92)**

B

**Solution to Exercise 5.17 (p. 92)**

C

**Solution to Exercise 5.18 (p. 92)**

D

**Solution to Exercise 5.19 (p. 92)**

B

**Solution to Exercise 5.20 (p. 92)**

B

**Solution to Exercise 5.21 (p. 93)****REVIEW QUESTION 1 Solution**

- a. 24%
- b. 27%

**Solution to Exercise 5.22 (p. 93)**

REVIEW QUESTION 2 Solution : Qualitative

**Solution to Exercise 5.23 (p. 93)**REVIEW QUESTION 3 Solution :  $P(\text{PT} | E)$ **Solution to Exercise 5.24 (p. 93)**

REVIEW QUESTION 4 Solution : 0.7336

**Solution to Exercise 5.25 (p. 93)****REVIEW QUESTION 5 Solution**

- a. No,
- b. No,

**Solution to Exercise 5.26 (p. 93)**REVIEW QUESTION 6 Solution :  $B(10, 0.76)$ **Solution to Exercise 5.27 (p. 93)**

REVIEW QUESTION 7 Solution : 0.9330

**Solution to Exercise 5.28 (p. 93)****REVIEW QUESTION 8 Solution**

- a.  $X$  = the number of questions posted to the Statistics Listserv per day
- b.  $x = 0, 1, 2, \dots$
- c.  $X \sim P(2)$
- d. 0

**Solution to Exercise 5.29 (p. 94)**

REVIEW QUESTION 9 Solution : \$150

**Solution to Exercise 5.30 (p. 94)**

REVIEW QUESTION 10 Solution : Matt

**Solution to Exercise 5.31 (p. 94)**

REVIEW QUESTION 11 Solution : Mean

**Solution to Exercise 5.32 (p. 94)****REVIEW QUESTION 12 Solution**

- a. False
- b. True
- c. False
- d. False

**Solution to Exercise 5.33 (p. 94)**

REVIEW QUESTION 13 Solution : 16

**Solution to Exercise 5.34 (p. 94)**

REVIEW QUESTION 14 Solution : 2, 2, 3

**Solution to Exercise 5.35 (p. 95)**REVIEW QUESTION 15 Solution :  $\frac{5}{10} = 0.5$ **Solution to Exercise 5.36 (p. 95)**REVIEW QUESTION 16 Solution :  $\frac{7}{15}$ **Solution to Exercise 5.37 (p. 95)**REVIEW QUESTION 17 Solution :  $\frac{2}{15}$

# Chapter 6

## The Normal Distribution

### 6.1 Summary of Formulas<sup>1</sup>

**Rule 6.1:** Normal Probability Distribution

$$X \sim N(\mu, \sigma)$$

$\mu$  = the mean       $\sigma$  = the standard deviation

**Rule 6.2:** Standard Normal Probability Distribution

$$Z \sim N(0, 1)$$

$Z$  = a standardized value (z-score)

mean = 0      standard deviation = 1

**Rule 6.3:** Finding the kth Percentile

To find the  $k$ th percentile when the z-score is known:  $k = \mu + (z) \sigma$

**Rule 6.4:** z-score

$$z = \frac{x - \mu}{\sigma}$$

**Rule 6.5:** Finding the area to the left

The area to the left:  $P(X < x)$

**Rule 6.6:** Finding the area to the right

The area to the right:  $P(X > x) = 1 - P(X < x)$

### 6.2 Homework<sup>2</sup>

#### Exercise 6.1

(*Solution on p. 107.*)

According to a study done by De Anza students, the height for Asian adult males is normally distributed with an average of 66 inches and a standard deviation of 2.5 inches. Suppose one Asian adult male is randomly chosen. Let  $X$  =height of the individual.

- a.  $X \sim \underline{\hspace{2cm}} (\underline{\hspace{1cm}}, \underline{\hspace{1cm}})$
- b. Find the probability that the person is between 65 and 69 inches. Include a sketch of the graph and write a probability statement.
- c. Would you expect to meet many Asian adult males over 72 inches? Explain why or why not, and justify your answer numerically.

<sup>1</sup>This content is available online at <<http://cnx.org/content/m16987/1.4/>>.

<sup>2</sup>This content is available online at <<http://cnx.org/content/m16978/1.20/>>.

- d. The middle 40% of heights fall between what two values? Sketch the graph and write the probability statement.

**Exercise 6.2**

IQ is normally distributed with a mean of 100 and a standard deviation of 15. Suppose one individual is randomly chosen. Let  $X = \text{IQ}$  of an individual.

- $X \sim \underline{\hspace{2cm}}(\underline{\hspace{2cm}}, \underline{\hspace{2cm}})$
- Find the probability that the person has an IQ greater than 120. Include a sketch of the graph and write a probability statement.
- Mensa is an organization whose members have the top 2% of all IQs. Find the minimum IQ needed to qualify for the Mensa organization. Sketch the graph and write the probability statement.
- The middle 50% of IQs fall between what two values? Sketch the graph and write the probability statement.

**Exercise 6.3**

(*Solution on p. 107.*)

The percent of fat calories that a person in America consumes each day is normally distributed with a mean of about 36 and a standard deviation of 10. Suppose that one individual is randomly chosen. Let  $X = \text{percent of fat calories}$ .

- $X \sim \underline{\hspace{2cm}}(\underline{\hspace{2cm}}, \underline{\hspace{2cm}})$
- Find the probability that the percent of fat calories a person consumes is more than 40. Graph the situation. Shade in the area to be determined.
- Find the maximum number for the lower quarter of percent of fat calories. Sketch the graph and write the probability statement.

**Exercise 6.4**

Suppose that the distance of fly balls hit to the outfield (in baseball) is normally distributed with a mean of 250 feet and a standard deviation of 50 feet.

- If  $X = \text{distance in feet for a fly ball}$ , then  $X \sim \underline{\hspace{2cm}}(\underline{\hspace{2cm}}, \underline{\hspace{2cm}})$
- If one fly ball is randomly chosen from this distribution, what is the probability that this ball traveled fewer than 220 feet? Sketch the graph. Scale the horizontal axis  $X$ . Shade the region corresponding to the probability. Find the probability.
- Find the 80th percentile of the distribution of fly balls. Sketch the graph and write the probability statement.

**Exercise 6.5**

(*Solution on p. 107.*)

In China, 4-year-olds average 3 hours a day unsupervised. Most of the unsupervised children live in rural areas, considered safe. Suppose that the standard deviation is 1.5 hours and the amount of time spent alone is normally distributed. We randomly survey one Chinese 4-year-old living in a rural area. We are interested in the amount of time the child spends alone per day. (Source: **San Jose Mercury News**)

- In words, define the random variable  $X$ .  $X =$
- $X \sim$
- Find the probability that the child spends less than 1 hour per day unsupervised. Sketch the graph and write the probability statement.
- What percent of the children spend over 10 hours per day unsupervised?
- 70% of the children spend at least how long per day unsupervised?

**Exercise 6.6**

In the 1992 presidential election, Alaska's 40 election districts averaged 1956.8 votes per district for President Clinton. The standard deviation was 572.3. (There are only 40 election districts in Alaska.) The distribution of the votes per district for President Clinton was bell-shaped. Let  $X$  = number of votes for President Clinton for an election district. (Source: **The World Almanac and Book of Facts**)

- State the approximate distribution of  $X$ .  $X \sim$
- Is 1956.8 a population mean or a sample mean? How do you know?
- Find the probability that a randomly selected district had fewer than 1600 votes for President Clinton. Sketch the graph and write the probability statement.
- Find the probability that a randomly selected district had between 1800 and 2000 votes for President Clinton.
- Find the third quartile for votes for President Clinton.

**Exercise 6.7**

(Solution on p. 107.)

Suppose that the duration of a particular type of criminal trial is known to be normally distributed with a mean of 21 days and a standard deviation of 7 days.

- In words, define the random variable  $X$ .  $X =$
- $X \sim$
- If one of the trials is randomly chosen, find the probability that it lasted at least 24 days. Sketch the graph and write the probability statement.
- 60% of all of these types of trials are completed within how many days?

**Exercise 6.8**

Terri Vogel, an amateur motorcycle racer, averages 129.71 seconds per 2.5 mile lap (in a 7 lap race) with a standard deviation of 2.28 seconds. The distribution of her race times is normally distributed. We are interested in one of her randomly selected laps. (Source: log book of Terri Vogel)

- In words, define the random variable  $X$ .  $X =$
- $X \sim$
- Find the percent of her laps that are completed in less than 130 seconds.
- The fastest 3% of her laps are under \_\_\_\_\_.
- The middle 80% of her laps are from \_\_\_\_\_ seconds to \_\_\_\_\_ seconds.

**Exercise 6.9**

(Solution on p. 107.)

Thuy Dau, Ngoc Bui, Sam Su, and Lan Young conducted a survey as to how long customers at Lucky claimed to wait in the checkout line until their turn. Let  $X$  = time in line. Below are the ordered real data (in minutes):

0.50	4.25	5	6	7.25
1.75	4.25	5.25	6	7.25
2	4.25	5.25	6.25	7.25
2.25	4.25	5.5	6.25	7.75
2.25	4.5	5.5	6.5	8
2.5	4.75	5.5	6.5	8.25
2.75	4.75	5.75	6.5	9.5
3.25	4.75	5.75	6.75	9.5
3.75	5	6	6.75	9.75
3.75	5	6	6.75	10.75

**Table 6.1**

- a. Calculate the sample mean and the sample standard deviation.
- b. Construct a histogram. Start the  $x$  – axis at  $-0.375$  and make bar widths of 2 minutes.
- c. Draw a smooth curve through the midpoints of the tops of the bars.
- d. In words, describe the shape of your histogram and smooth curve.
- e. Let the sample mean approximate  $\mu$  and the sample standard deviation approximate  $\sigma$ . The distribution of  $X$  can then be approximated by  $X \sim$
- f. Use the distribution in (e) to calculate the probability that a person will wait fewer than 6.1 minutes.
- g. Determine the cumulative relative frequency for waiting less than 6.1 minutes.
- h. Why aren't the answers to (f) and (g) exactly the same?
- i. Why are the answers to (f) and (g) as close as they are?
- j. If only 10 customers were surveyed instead of 50, do you think the answers to (f) and (g) would have been closer together or farther apart? Explain your conclusion.

**Exercise 6.10**

Suppose that Ricardo and Anita attend different colleges. Ricardo's GPA is the same as the average GPA at his school. Anita's GPA is 0.70 standard deviations above her school average. In complete sentences, explain why each of the following statements may be false.

- a. Ricardo's actual GPA is lower than Anita's actual GPA.
- b. Ricardo is not passing since his z-score is zero.
- c. Anita is in the 70th percentile of students at her college.

**Exercise 6.11**

(Solution on p. 107.)

Below is a sample of the maximum capacity (maximum number of spectators) of sports stadiums. The table does not include horse racing or motor racing stadiums. (Source: [http://en.wikipedia.org/wiki/List\\_of\\_stadiums\\_by\\_capacity](http://en.wikipedia.org/wiki/List_of_stadiums_by_capacity))

40,000	40,000	45,050	45,500	46,249	48,134
49,133	50,071	50,096	50,466	50,832	51,100
51,500	51,900	52,000	52,132	52,200	52,530
52,692	53,864	54,000	55,000	55,000	55,000
55,000	55,000	55,000	55,082	57,000	58,008
59,680	60,000	60,000	60,492	60,580	62,380
62,872	64,035	65,000	65,050	65,647	66,000
66,161	67,428	68,349	68,976	69,372	70,107
70,585	71,594	72,000	72,922	73,379	74,500
75,025	76,212	78,000	80,000	80,000	82,300

**Table 6.2**

- a. Calculate the sample mean and the sample standard deviation for the maximum capacity of sports stadiums (the data).
- b. Construct a histogram of the data.
- c. Draw a smooth curve through the midpoints of the tops of the bars of the histogram.
- d. In words, describe the shape of your histogram and smooth curve.
- e. Let the sample mean approximate  $\mu$  and the sample standard deviation approximate  $\sigma$ . The distribution of  $X$  can then be approximated by  $X \sim$
- f. Use the distribution in (e) to calculate the probability that the maximum capacity of sports stadiums is less than 67,000 spectators.
- g. Determine the cumulative relative frequency that the maximum capacity of sports stadiums is less than 67,000 spectators. Hint: Order the data and count the sports stadiums that have a maximum capacity less than 67,000. Divide by the total number of sports stadiums in the sample.
- h. Why aren't the answers to (f) and (g) exactly the same?

### 6.2.1 Try These Multiple Choice Questions

**The questions below refer to the following:** The patient recovery time from a particular surgical procedure is normally distributed with a mean of 5.3 days and a standard deviation of 2.1 days.

**Exercise 6.12**

What is the median recovery time?

- A. 2.7
- B. 5.3
- C. 7.4
- D. 2.1

(*Solution on p. 107.*)

**Exercise 6.13**

(*Solution on p. 107.*)

What is the z-score for a patient who takes 10 days to recover?

- A. 1.5
- B. 0.2

- C. 2.2  
D. 7.3

**Exercise 6.14**

(Solution on p. 107.)

What is the probability of spending more than 2 days in recovery?

- A. 0.0580  
B. 0.8447  
C. 0.0553  
D. 0.9420

**Exercise 6.15**

(Solution on p. 107.)

The 90th percentile for recovery times is?

- A. 8.89  
B. 7.07  
C. 7.99  
D. 4.32

**The questions below refer to the following:** The length of time to find a parking space at 9 A.M. follows a normal distribution with a mean of 5 minutes and a standard deviation of 2 minutes.

**Exercise 6.16**

(Solution on p. 107.)

Based upon the above information and numerically justified, would you be surprised if it took less than 1 minute to find a parking space?

- A. Yes  
B. No  
C. Unable to determine

**Exercise 6.17**

(Solution on p. 107.)

Find the probability that it takes at least 8 minutes to find a parking space.

- A. 0.0001  
B. 0.9270  
C. 0.1862  
D. 0.0668

**Exercise 6.18**

(Solution on p. 107.)

Seventy percent of the time, it takes more than how many minutes to find a parking space?

- A. 1.24  
B. 2.41  
C. 3.95  
D. 6.05

**Exercise 6.19**

(Solution on p. 108.)

If the mean is significantly greater than the standard deviation, which of the following statements is true?

- I. The data cannot follow the uniform distribution.  
II. The data cannot follow the exponential distribution..  
III. The data cannot follow the normal distribution.

- A. I only
- B. II only
- C. III only
- D. I, II, and III

### 6.3 Review Questions<sup>3</sup>

The next two questions refer to:  $X \sim U(3, 13)$

**Exercise 6.20: REVIEW QUESTION 1**

(Solution on p. 108.)

Explain which of the following are false and which are true.

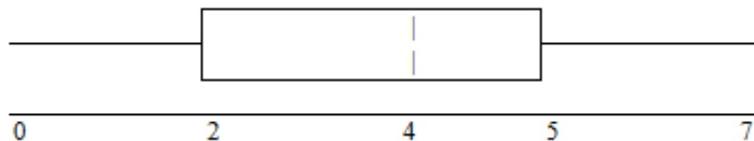
- a -  $f(x) = \frac{1}{10}, 3 \leq x \leq 13$
- b - There is no mode.
- c - The median is less than the mean.
- d -  $P(X > 10) = P(X \leq 6)$

**Exercise 6.21: REVIEW QUESTION 2**

(Solution on p. 108.)

Calculate:

- a - Mean
- b - Median
- c - 65th percentile.



**Exercise 6.22: REVIEW QUESTION 3**

(Solution on p. 108.)

Which of the following is true for the above box plot?

- a - 25% of the data are at most 5.
- b - There is about the same amount of data from 4 – 5 as there is from 5 – 7.
- c - There are no data values of 3.
- d - 50% of the data are 4.

**Exercise 6.23: REVIEW QUESTION 4**

(Solution on p. 108.)

If  $P(G | H) = P(G)$ , then which of the following is correct?

- A -  $G$  and  $H$  are mutually exclusive events.
- B -  $P(G) = P(H)$
- C - Knowing that  $H$  has occurred will affect the chance that  $G$  will happen.
- D -  $G$  and  $H$  are independent events.

**Exercise 6.24: REVIEW QUESTION 5**

(Solution on p. 108.)

If  $P(J) = 0.3$ ,  $P(K) = 0.6$ , and  $J$  and  $K$  are independent events, then explain which are correct and which are incorrect.

<sup>3</sup>This content is available online at <<http://cnx.org/content/m19027/1.1/>>.

- A -  $P(J \text{ and } K = 0)$
- B -  $P(J \text{ or } K = 0.9)$
- C -  $P(J \text{ or } K = 0.72)$
- D -  $P(J) \neq P(J | K)$

**Exercise 6.25: REVIEW QUESTION 6**

(Solution on p. 108.)

On average, 5 students from each high school class get full scholarships to 4-year colleges. Assume that most high school classes have about 500 students.

$X$  = the number of students from a high school class that get full scholarships to 4-year school.  
Which of the following is the distribution of  $X$ ?

- A.  $P(5)$
- B.  $B(500, 5)$
- C.  $\text{Exp}(1/5)$
- D.  $N(5, (0.01)(0.99)/500)$

## Solutions to Exercises in Chapter 6

### Solution to Exercise 6.1 (p. 99)

- a.  $N(66, 2.5)$
- b. 0.5404
- c. No
- d. Between 64.7 and 67.3 inches

### Solution to Exercise 6.3 (p. 100)

- a.  $N(36, 10)$
- b. 0.3446
- c. 29.3

### Solution to Exercise 6.5 (p. 100)

- a. the time (in hours) a 4-year-old in China spends unsupervised per day
- b.  $N(3, 1.5)$
- c. 0.0912
- d. 0
- e. 2.21 hours

### Solution to Exercise 6.7 (p. 101)

- a. The duration of a criminal trial
- b.  $N(21, 7)$
- c. 0.3341
- d. 22.77

### Solution to Exercise 6.9 (p. 101)

- a. The sample mean is 5.51 and the sample standard deviation is 2.15
- e.  $N(5.51, 2.15)$
- f. 0.6081
- g. 0.64

### Solution to Exercise 6.11 (p. 102)

- a. The sample mean is 60,136.4 and the sample standard deviation is 10,468.1.
- e.  $N(60136.4, 10468.1)$
- f. 0.7440
- g. 0.7167

### Solution to Exercise 6.12 (p. 103)

B

### Solution to Exercise 6.13 (p. 103)

C

### Solution to Exercise 6.14 (p. 104)

D

### Solution to Exercise 6.15 (p. 104)

C

### Solution to Exercise 6.16 (p. 104)

A

### Solution to Exercise 6.17 (p. 104)

D

**Solution to Exercise 6.18 (p. 104)**

C

**Solution to Exercise 6.19 (p. 104)**

B

**Solution to Exercise 6.20 (p. 105)**

**REVIEW QUESTION 1 Solution**

- a - True
- b - True
- c - False – the median and the mean are the same for this symmetric distribution
- d - True

**Solution to Exercise 6.21 (p. 105)**

**REVIEW QUESTION 2 Solution**

- a - 8
- b - 8
- c -  $P(X < k) = 0.65 = (k - 3) * \left(\frac{1}{10}\right)$ .  $k = 9.5$

**Solution to Exercise 6.22 (p. 105)**

**REVIEW QUESTION 3 Solution**

- a - False –  $\frac{3}{4}$  of the data are at most 5
- b - True – each quartile has 25% of the data
- c - False – that is unknown
- d - False – 50% of the data are 4 or less

**Solution to Exercise 6.23 (p. 105)**

**REVIEW QUESTION 4 Solution**

D

**Solution to Exercise 6.24 (p. 105)**

**REVIEW QUESTION 5 Solution**

- A - False – J and K are independent, so they are not mutually exclusive which would imply dependency
- B - False
- C - True – since  $P(J \text{ and } K \neq 0) = P(J)P(K) = 0.09$
- D - False – P(J and K ≠ 0) are independent which implies  $P(J) = P(J | K)$

**Solution to Exercise 6.25 (p. 106)**

**REVIEW QUESTION 6 Solution**

A

# Chapter 7

## The Central Limit Theorem

### 7.1 Summary of Formulas<sup>1</sup>

**Rule 7.1:** Central Limit Theorem for Sample Means (Averages)

$$\bar{X} \sim N\left(\mu_X, \frac{\sigma_X}{\sqrt{n}}\right) \quad \text{Mean for Averages } (\bar{X}): \quad \mu_X$$

**Rule 7.2:** Central Limit Theorem for Sample Means (Averages) Z-Score and Standard Error of the Mean

$$z = \frac{\bar{x} - \mu_X}{\left(\frac{\sigma_X}{\sqrt{n}}\right)} \quad \text{Standard Error of the Mean (Standard Deviation for Averages } (\bar{X})) : \quad \frac{\sigma_X}{\sqrt{n}}$$

**Rule 7.3:** Central Limit Theorem for Sums

$$\Sigma X \sim N[(n) \cdot \mu_X, \sqrt{n} \cdot \sigma_X] \quad \text{Mean for Sums } (\Sigma X): \quad n \cdot \mu_X$$

**Rule 7.4:** Central Limit Theorem for Sums Z-Score and Standard Deviation for Sums

$$z = \frac{\Sigma x - n \cdot \mu_X}{\sqrt{n} \cdot \sigma_X} \quad \text{Standard Deviation for Sums } (\Sigma X): \quad \sqrt{n} \cdot \sigma_X$$

### 7.2 Homework: CLT (modified R. Bloom)<sup>2</sup>

#### Exercise 7.1

(Solution on p. 118.)

$X \sim N(60, 9)$ . Suppose that you form random samples of 25 from this distribution. Let  $\bar{X}$  be the random variable of averages. For c - f, sketch the graph, shade the region, label and scale the horizontal axis for  $\bar{X}$ , and find the probability.

- a. Sketch the distributions of  $X$  and  $\bar{X}$  on the same graph.
- b.  $\bar{X} \sim$
- c.  $P(\bar{X} < 60) =$
- d. Find the 30th percentile.
- e.  $P(56 < \bar{X} < 62) =$
- f.  $P(18 < \bar{X} < 58) =$
- g. Find the minimum value for the upper quartile.

#### Exercise 7.2

Determine which of the following are true and which are false. Then, in complete sentences, justify your answers.

<sup>1</sup>This content is available online at <<http://cnx.org/content/m16956/1.6/>>.

<sup>2</sup>This content is available online at <<http://cnx.org/content/m18940/1.2/>>.

- When the sample size is large, the mean of  $\bar{X}$  is approximately equal to the mean of  $X$ .
- When the sample size is large,  $\bar{X}$  is approximately normally distributed.
- When the sample size is large, the standard deviation of  $\bar{X}$  is approximately the same as the standard deviation of  $X$ .

**Exercise 7.3**

(Solution on p. 118.)

The percent of fat calories that a person in America consumes each day is normally distributed with a mean of about 36 and a standard deviation of about 10. Suppose that 16 individuals are randomly chosen.

Let  $\bar{X}$  = average percent of fat calories.

- $\bar{X} \sim \text{_____} (\text{_____}, \text{_____})$
- For the group of 16, find the probability that the average percent of fat calories consumed is more than 5. Graph the situation and shade in the area to be determined.
- Find the first quartile for the average percent of fat calories.

**Exercise 7.4**

Previously, De Anza statistics students estimated that the amount of change daytime statistics students carry is exponentially distributed with a mean of \$0.88. Suppose that we randomly pick 25 daytime statistics students.

- In words,  $X$  =
- $X \sim$
- In words,  $\bar{X}$  =
- $\bar{X} \sim \text{_____} (\text{_____}, \text{_____})$
- Find the probability that an individual had between \$0.80 and \$1.00. Graph the situation and shade in the area to be determined.
- Find the probability that the average of the 25 students was between \$0.80 and \$1.00. Graph the situation and shade in the area to be determined.
- Explain the why there is a difference in (e) and (f).

**Exercise 7.5**

(Solution on p. 118.)

Suppose that the distance of fly balls hit to the outfield (in baseball) is normally distributed with a mean of 250 feet and a standard deviation of 50 feet. We randomly sample 49 fly balls.

- If  $\bar{X}$  = average distance in feet for 49 fly balls, then  $\bar{X} \sim \text{_____} (\text{_____}, \text{_____})$
- What is the probability that the 49 balls traveled an average of less than 240 feet? Sketch the graph. Scale the horizontal axis for  $\bar{X}$ . Shade the region corresponding to the probability. Find the probability.
- Find the 80th percentile of the distribution of the average of 49 fly balls.

**Exercise 7.6**

Question removed from textbook.

**Exercise 7.7**

(Solution on p. 118.)

Note: Problem has been changed from original version of textbook.

Suppose that the duration of a particular type of criminal trial is known to have a mean of 21 days and a standard deviation of 7 days. We randomly sample 25 trials.

- Find the probability that the average length of the 25 trials is at least 24 days.
- Find the 10th percentile for the average length for samples of 25 trials of this type.

**Exercise 7.8**

According to the Internal Revenue Service, the average length of time for an individual to complete (record keep, learn, prepare, copy, assemble and send) IRS Form 1040 is 10.53 hours (without any attached schedules). The distribution is unknown. Let us assume that the standard deviation is 2 hours. Suppose we randomly sample 36 taxpayers.

- In words,  $X =$
- In words,  $\bar{X} =$
- $\bar{X} \sim$
- Would you be surprised if the 36 taxpayers finished their Form 1040s in an average of more than 12 hours? Explain why or why not in complete sentences.
- Would you be surprised if one taxpayer finished his Form 1040 in more than 12 hours? In a complete sentence, explain why.

**Exercise 7.9***(Solution on p. 118.)*

Suppose that a category of world class runners are known to run a marathon (26 miles) in an average of 145 minutes with a standard deviation of 14 minutes. Consider 49 of the races.

Let  $\bar{X}$  = the average of the 49 races.

- $\bar{X} \sim$
- Find the probability that the runner will average between 142 and 146 minutes in these 49 marathons.
- Find the 80th percentile for the average of these 49 marathons.
- Find the median of the average running times.

**Exercise 7.10**

The attention span of a two year-old is exponentially distributed with a mean of about 8 minutes. Suppose we randomly survey 60 two year-olds.

- In words,  $X =$
- $X \sim$
- In words,  $\bar{X} =$
- $\bar{X} \sim$
- Before doing any calculations, which do you think will be higher? Explain why.
  - the probability that an individual attention span is less than 10 minutes; or
  - the probability that the average attention span for the 60 children is less than 10 minutes?

Why?
- Calculate the probabilities in part (e).
- Explain why the distribution for  $\bar{X}$  is not exponential.

**Exercise 7.11***(Solution on p. 118.)*

Note: Parts g,h,i,j of this problem have been changed from original version of textbook.

Suppose that the length of research papers is uniformly distributed from 10 to 25 pages. (Use the continuous uniform distribution - assume that the page count measures fractional pages.) We survey a random sample of 55 research papers turned in to a professor. We are interested in the average length of the research papers.

- In words,  $X =$
- $X \sim$
- $\mu_X =$

- d.  $\sigma_X =$
- e. In words,  $\bar{X} =$
- f.  $\bar{X} \sim$
- g. Find the probability that an individual paper is longer than 18 pages.
- h. Find the probability that the average length of the 55 papers is more than 18 pages.
- i. Find the 64th percentile for the length of individual papers.
- j. Find the 64th percentile for the average length for samples of papers.
- k. Why is it so unlikely that the average length of the papers will be less than 12 pages?

**Exercise 7.12**

The length of songs in a collector's CD collection is uniformly distributed from 2 to 3.5 minutes. Suppose we randomly pick 5 CDs from the collection. There is a total of 43 songs on the 5 CDs.

- a. In words,  $X =$
- b.  $X \sim$
- c. In words,  $\bar{X} =$
- d.  $\bar{X} \sim$
- e. Find the first quartile for the average song length.
- f. The IQR (interquartile range) for the average song length is from \_\_\_\_\_ to \_\_\_\_\_.

**Exercise 7.13**

*(Solution on p. 118.)*

Note: Parts d,e,f of this problem have been changed from original version of textbook.

Salaries for teachers in a particular elementary school district are normally distributed with a mean of \$44,000 and a standard deviation of \$6500. We randomly survey 10 teachers from that district.

- a. In words,  $X =$
- b. In words,  $\bar{X} =$
- c.  $\bar{X} \sim$
- d. Find the probability that an individual teacher earns more than \$40,000.
- e. Find the probability that the average salary for the sample is more than \$40,000.
- f. Find the probability that the average salary for the sample is more than \$50,000.
- g. Find the 90th percentile for an individual teacher's salary.
- h. Find the 90th percentile for the average teachers' salary for samples of 10 teachers.
- i. If we surveyed 70 teachers instead of 10, graphically, how would that change the distribution for  $\bar{X}$ ?
- j. If each of the teachers in this elementary school district received a \$3000 raise, graphically, how would that change the distribution for  $\bar{X}$ ?

**Exercise 7.14**

The distribution of income in some Third World countries is considered wedge shaped (many very poor people, very few middle income people, and few to many wealthy people). Suppose we pick a country with a wedge distribution. Let the average salary be \$2000 per year with a standard deviation of \$8000. We randomly survey 1000 residents of that country.

- a. In words,  $X =$
- b. In words,  $\bar{X} =$
- c.  $\bar{X} \sim$
- d. How is it possible for the standard deviation to be greater than the average?
- e. Why is it more likely that the average of the 1000 residents will be from \$2000 to \$2100 than from \$2100 to \$2200?

**Exercise 7.15**

(Solution on p. 119.)

The average length of a maternity stay in a U.S. hospital is said to be 2.4 days with a standard deviation of 0.9 days. We randomly survey 80 women who recently bore children in a U.S. hospital.

- In words,  $X$  =
- In words,  $\bar{X}$  =
- $\bar{X} \sim$
- Question removed from text
- Question removed from text
- Is it likely that an individual stayed more than 5 days in the hospital? Why or why not?
- Is it likely that the average stay for the 80 women was more than 5 days? Why or why not?
- Which is more likely:
  - an individual stayed more than 5 days; or
  - the average stay of 80 women was more than 5 days?

**Exercise 7.16**

In 1940 the average size of a U.S. farm was 174 acres. Let's say that the standard deviation was 55 acres. Suppose we randomly survey 38 farmers from 1940. (Source: U.S. Dept. of Agriculture)

- In words,  $X$  =
- In words,  $\bar{X}$  =
- $\bar{X} \sim$
- The IQR for  $\bar{X}$  is from \_\_\_\_\_ acres to \_\_\_\_\_ acres.

**Exercise 7.17**

(Solution on p. 119.)

The stock closing prices of 35 U.S. semiconductor manufacturers are given below. (Source: **Wall Street Journal**)

8.625; 30.25; 27.625; 46.75; 32.875; 18.25; 5; 0.125; 2.9375; 6.875; 28.25; 24.25; 21; 1.5; 30.25; 71; 43.5; 49.25; 2.5625; 31; 16.5; 9.5; 18.5; 18; 9; 10.5; 16.625; 1.25; 18; 12.875; 7; 2.875; 2.875; 60.25; 29.25

- In words,  $X$  =
- i.  $\bar{x}$  =  
ii.  $s_x$  =  
iii.  $n$  =
- Construct a histogram of the distribution of the averages. Start at  $x = -0.0005$ . Make bar widths of 10.
- In words, describe the distribution of stock prices.
- Randomly average 5 stock prices together. (Use a random number generator.) Continue averaging 5 pieces together until you have 10 averages. List those 10 averages.
- Use the 10 averages from (e) to calculate:
  - $\bar{x}$  =
  - $s_{\bar{x}}$  =
- Construct a histogram of the distribution of the averages. Start at  $x = -0.0005$ . Make bar widths of 10.
- Does this histogram look like the graph in (c)?
- In 1 - 2 complete sentences, explain why the graphs either look the same or look different?
- Based upon the theory of the Central Limit Theorem,  $\bar{X} \sim$

**Exercise 7.18**

Use the Initial Public Offering data (Section 14.1.2: Stock Prices) (see "Table of Contents") to do this problem.

- a. In words,  $X =$
- b. i.  $\mu_X =$   
ii.  $\sigma_X =$   
iii.  $n =$
- c. Construct a histogram of the distribution. Start at  $x = -0.50$ . Make bar widths of \$5.
- d. In words, describe the distribution of stock prices.
- e. Randomly average 5 stock prices together. (Use a random number generator.) Continue averaging 5 pieces together until you have 15 averages. List those 15 averages.
- f. Use the 15 averages from (e) to calculate the following:
  - i.  $\bar{x} =$
  - ii.  $\bar{s}_x =$
- g. Construct a histogram of the distribution of the averages. Start at  $x = -0.50$ . Make bar widths of \$5.
- h. Does this histogram look like the graph in (c)? Explain any differences.
- i. In 1 - 2 complete sentences, explain why the graphs either look the same or look different?
- j. Based upon the theory of the Central Limit Theorem,  $\bar{X} \sim$

### 7.2.1 Try these multiple choice questions.

**The next two questions refer to the following information:** The time to wait for a particular rural bus is distributed uniformly from 0 to 75 minutes. 100 riders are randomly sampled to learn how long they waited.

**Exercise 7.19**

(*Solution on p. 119.*)

The 90th percentile sample average wait time (in minutes) for a sample of 100 riders is:

- A. 315.0
- B. 40.3
- C. 38.5
- D. 65.2

**Exercise 7.20**

(*Solution on p. 119.*)

Would you be surprised, based upon numerical calculations, if the sample average wait time (in minutes) for 100 riders was less than 30 minutes?

- A. Yes
- B. No
- C. There is not enough information.

**Exercise 7.21**

(*Solution on p. 119.*)

Which of the following is NOT TRUE about the distribution for averages?

- A. The mean, median and mode are equal
- B. The area under the curve is one
- C. The curve never touches the x-axis
- D. The curve is skewed to the right

**The next three questions refer to the following information:** The cost of unleaded gasoline in the Bay Area once followed an unknown distribution with a mean of \$2.59 and a standard deviation of \$0.10. Thirty gas stations from the Bay Area are randomly chosen. We are interested in the average cost of gasoline for the 30 gas stations.

NOTE: The situation for problems 22 and 23 has been changed from the original version of the textbook.

**Exercise 7.22**

(*Solution on p. 119.*)

The distribution to use for the average cost of gasoline for the 30 gas stations is

- A.  $\bar{X} \sim N(2.59, 0.10)$
- B.  $\bar{X} \sim N\left(2.59, \frac{0.10}{\sqrt{30}}\right)$
- C.  $\bar{X} \sim N\left(2.59, \frac{0.10}{30}\right)$
- D.  $\bar{X} \sim N\left(2.59, \frac{30}{0.10}\right)$

**Exercise 7.23**

(*Solution on p. 119.*)

What is the probability that the average price for 30 gas stations is over \$2.69?

- A. Almost zero
- B. 0.1587
- C. 0.0943
- D. Unknown

**Exercise 7.24**

(*Solution on p. 119.*)

Find the probability that the average price for 30 gas stations is less than \$2.55.

- A. 0.6554
- B. 0.3446
- C. 0.0142
- D. 0.9858
- E. 0

### 7.3 Review Questions<sup>3</sup>

**The next three questions refer to the following information:** Richard's Furniture Company delivers furniture from 10 A.M. to 2 P.M. continuously and uniformly. We are interested in how long (in hours) past the 10 A.M. start time that individuals wait for their delivery.

**Exercise 7.25: REVIEW QUESTION 1**

(*Solution on p. 119.*)

$X \sim$

- A.  $U(0, 4)$
- B.  $U(10, 2)$
- C.  $Exp(2)$
- D.  $N(2, 1)$

**Exercise 7.26: REVIEW QUESTION 2**

(*Solution on p. 119.*)

The average wait time is:

- A. 1 hour
- B. 2 hour
- C. 2.5 hour

---

<sup>3</sup>This content is available online at <<http://cnx.org/content/m18863/1.2/>>.

D. 4 hour

**Exercise 7.27: REVIEW QUESTION 3**

(*Solution on p. 119.*)

Suppose that it is now past noon on a delivery day. The probability that a person must wait at least  $1\frac{1}{2}$  more hours is:

- A.  $\frac{1}{4}$
- B.  $\frac{1}{2}$
- C.  $\frac{3}{4}$
- D.  $\frac{3}{8}$

**Exercise 7.28: REVIEW QUESTION 4**

(*Solution on p. 119.*)

Given:  $X \sim \text{Exp}\left(\frac{1}{3}\right)$ .

- a. Find  $P(X > 1)$
- b. Calculate the minimum value for the upper quartile.
- c. Find  $P\left(X = \frac{1}{3}\right)$

**Exercise 7.29: REVIEW QUESTION 5**

(*Solution on p. 119.*)

- 40% of full-time students took 4 years to graduate
- 30% of full-time students took 5 years to graduate
- 20% of full-time students took 6 years to graduate
- 10% of full-time students took 7 years to graduate

The expected time for full-time students to graduate is:

- A. 4 years
- B. 4.5 years
- C. 5 years
- D. 5.5 years

**Exercise 7.30: REVIEW QUESTION 6**

(*Solution on p. 119.*)

Which of the following distributions is described by the following example?

Many people can run a short distance of under 2 miles, but as the distance increases, fewer people can run that far.

- A. Binomial
- B. Uniform
- C. Exponential
- D. Normal

**Exercise 7.31: REVIEW QUESTION 7**

(*Solution on p. 119.*)

The length of time to brush one's teeth is generally thought to be exponentially distributed with a mean of  $\frac{3}{4}$  minutes. Find the probability that a randomly selected person brushes his/her teeth less than  $\frac{3}{4}$  minutes.

- A. 0.5
- B.  $\frac{3}{4}$
- C. 0.43
- D. 0.63

**Exercise 7.32: REVIEW QUESTION 8***(Solution on p. 119.)***Which distribution accurately describes the following situation?**

A 2008 report on technology use states that approximately 20 percent of U.S. households have never sent an e-mail. (source: <http://www.webguild.org/2008/05/20-percent-of-americans-have-never-used-email.php>) Suppose that we select a random sample of fourteen U.S. households

X =the number of households in the sample of 14 households that have never sent an email

- A.  $B(14, 0.20)$
- B.  $P(2.8)$
- C.  $N(2.8, 2.24)$
- D.  $\text{Exp}\left(\frac{1}{0.20}\right)$

## Solutions to Exercises in Chapter 7

### Solution to Exercise 7.1 (p. 109)

- b.  $X\bar{\sim}N\left(60, \frac{9}{\sqrt{25}}\right)$
- c. 0.5000
- d. 59.06
- e. 0.8536
- f. 0.1333
- h. 61.214

### Solution to Exercise 7.3 (p. 110)

- a.  $N\left(36, \frac{10}{\sqrt{16}}\right)$
- b. 1
- c. 34.31

### Solution to Exercise 7.5 (p. 110)

- a.  $N\left(250, \frac{50}{\sqrt{49}}\right)$
- b. 0.0808
- c. 256.01 feet

### Solution to Exercise 7.7 (p. 110)

- a: Answer: 0.0161
- b: Answer: 19.2 days

### Solution to Exercise 7.9 (p. 111)

- a.  $N\left(145, \frac{14}{\sqrt{49}}\right)$
- b. 0.6247
- c. 146.68 minutes
- d. 145 minutes

### Solution to Exercise 7.11 (p. 111)

- a.  $X$  = the number of pages in one individual paper
- b.  $U(10, 25)$
- c. 17.5
- d. 4.33
- e.  $\bar{X}$  = the average length of papers in sample of 55 papers
- f.  $N(17.5, 0.5839)$
- g.  $P(X > 18) = 0.4667$
- h.  $P(\bar{X} > 18) = 0.1959$
- i. 19.6 pages: 64% of papers are less than 19.6 pages long
- j. 17.7 pages: 64% of samples of 55 papers have average length less than 17.7 pages.

### Solution to Exercise 7.13 (p. 112)

- a.  $X$  = the salary earned by one individual teacher
- b.  $\bar{X}$  = the average salary for the 10 teachers in the sample
- c.  $N\left(44,000, \frac{6500}{\sqrt{10}}\right)$
- d. 0.7308

- e. 0.9742
- f. 0.0018
- g. \$52,330
- h. \$46,634
- i. The distribution would be more concentrated about the mean. According to the CLT, the spread of the distribution for the sample mean gets smaller when the sample size is increased.
- j. If each teacher received a \$3,000 raise, then the population mean would increase by \$3,000. The population mean is at the center of the distribution for the sample mean. So the distribution for the sample mean would shift \$3,000 to the right along the horizontal axis.

**Solution to Exercise 7.15 (p. 113)**

- c.  $N\left(2.4, \frac{0.9}{\sqrt{80}}\right)$
- h. Individual

**Solution to Exercise 7.17 (p. 113)**

- b. \$20.71; \$17.31; 35
- d. Exponential distribution,  $X \sim Exp(1/20.71)$
- f. \$20.71; \$11.14
- j.  $N\left(20.71, \frac{17.31}{\sqrt{5}}\right)$

**Solution to Exercise 7.19 (p. 114)**

B

**Solution to Exercise 7.20 (p. 114)**

A

**Solution to Exercise 7.21 (p. 114)**

D

**Solution to Exercise 7.22 (p. 115)**

B

**Solution to Exercise 7.23 (p. 115)**

A

**Solution to Exercise 7.24 (p. 115)**

C

**Solution to Exercise 7.25 (p. 115)**

REVIEW QUESTION 1 Solution : A

**Solution to Exercise 7.26 (p. 115)**

REVIEW QUESTION 1 Solution : B

**Solution to Exercise 7.27 (p. 116)**

REVIEW QUESTION 3 Solution : A

**Solution to Exercise 7.28 (p. 116)****REVIEW QUESTION 4 Solution**

- a. 0.7165
- b. 4.16
- c. 0

**Solution to Exercise 7.29 (p. 116)**

REVIEW QUESTION 5 Solution : C

**Solution to Exercise 7.30 (p. 116)**

REVIEW QUESTION 6 Solution : C

**Solution to Exercise 7.31 (p. 116)**

REVIEW QUESTION 7 Solution : D

**Solution to Exercise 7.32 (p. 117)**

REVIEW QUESTION 8 Solution : A



# Chapter 8

## Confidence Intervals

### 8.1 Summary of Formulas<sup>1</sup>

**Formula 8.1:** General form of a confidence interval

(lower value, upper value) = (point estimate – error bound, point estimate + error bound)

**Formula 8.2:** To find the error bound when you know the confidence interval

error bound = upper value – point estimate      OR      error bound =  $\frac{\text{upper value} - \text{lower value}}{2}$

**Formula 8.3:** Single Population Mean, Known Standard Deviation, Normal Distribution

Use the Normal Distribution for Means<sup>2</sup>      EBM =  $z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

The confidence interval has the format  $(\bar{x} - \text{EBM}, \bar{x} + \text{EBM})$ .

**Formula 8.4:** Single Population Mean, Unknown Standard Deviation, Student-t Distribution

Use the Student-t Distribution with degrees of freedom  $df = n - 1$ . EBM =  $t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$

**Formula 8.5:** Single Population Proportion, Normal Distribution

Use the Normal Distribution for a single population proportion  $p' = \frac{x}{n}$

EBP =  $z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p' \cdot q'}{n}}$        $p' + q' = 1$

The confidence interval has the format  $(p' - \text{EBP}, p' + \text{EBP})$ .

**Formula 8.6:** Point Estimates

$\bar{x}$  is a point estimate for  $\mu$

$p'$  is a point estimate for  $\rho$

$s$  is a point estimate for  $\sigma$

### 8.2 Homework<sup>3</sup>

NOTE: If you are using a student-t distribution for a homework problem below, you may assume that the underlying population is normally distributed. (In general, you must first prove that assumption, though.)

#### Exercise 8.1

(*Solution on p. 135.*)

Among various ethnic groups, the standard deviation of heights is known to be approximately 3 inches. We wish to construct a 95% confidence interval for the average height of male Swedes. 48

<sup>1</sup>This content is available online at <<http://cnx.org/content/m16973/1.7/>>.

<sup>2</sup>"Central Limit Theorem: Central Limit Theorem for Sample Means (Averages)" <<http://cnx.org/content/m16947/latest/>>

<sup>3</sup>This content is available online at <<http://cnx.org/content/m16966/1.12/>>.

male Swedes are surveyed. The sample mean is 71 inches. The sample standard deviation is 2.8 inches.

- a. i.  $\bar{x} = \underline{\hspace{2cm}}$
- ii.  $\sigma = \underline{\hspace{2cm}}$
- iii.  $s_x = \underline{\hspace{2cm}}$
- iv.  $n = \underline{\hspace{2cm}}$
- v.  $n - 1 = \underline{\hspace{2cm}}$
- b. Define the Random Variables  $X$  and  $\bar{X}$ , in words.
- c. Which distribution should you use for this problem? Explain your choice.
- d. Construct a 95% confidence interval for the population average height of male Swedes.
  - i. State the confidence interval.
  - ii. Sketch the graph.
  - iii. Calculate the error bound.
- e. What will happen to the level of confidence obtained if 1000 male Swedes are surveyed instead of 48? Why?

### Exercise 8.2

In six packages of “The Flintstones® Real Fruit Snacks” there were 5 Bam-Bam snack pieces. The total number of snack pieces in the six bags was 68. We wish to calculate a 96% confidence interval for the population proportion of Bam-Bam snack pieces.

- a. Define the Random Variables  $X$  and  $P'$ , in words.
- b. Which distribution should you use for this problem? Explain your choice
- c. Calculate  $p'$ .
- d. Construct a 96% confidence interval for the population proportion of Bam-Bam snack pieces per bag.
  - i. State the confidence interval.
  - ii. Sketch the graph.
  - iii. Calculate the error bound.
- e. Do you think that six packages of fruit snacks yield enough data to give accurate results? Why or why not?

### Exercise 8.3

(*Solution on p. 135.*)

A random survey of enrollment at 35 community colleges across the United States yielded the following figures (source: Microsoft Bookshelf): 6414; 1550; 2109; 9350; 21828; 4300; 5944; 5722; 2825; 2044; 5481; 5200; 5853; 2750; 10012; 6357; 27000; 9414; 7681; 3200; 17500; 9200; 7380; 18314; 6557; 13713; 17768; 7493; 2771; 2861; 1263; 7285; 28165; 5080; 11622. Assume the underlying population is normal.

- a. i.  $\bar{x} = \underline{\hspace{2cm}}$
- ii.  $s_x = \underline{\hspace{2cm}}$
- iii.  $n = \underline{\hspace{2cm}}$
- iv.  $n - 1 = \underline{\hspace{2cm}}$
- b. Define the Random Variables  $X$  and  $\bar{X}$ , in words.
- c. Which distribution should you use for this problem? Explain your choice.
- d. Construct a 95% confidence interval for the population average enrollment at community colleges in the United States.
  - i. State the confidence interval.
  - ii. Sketch the graph.
  - iii. Calculate the error bound.

- e. What will happen to the error bound and confidence interval if 500 community colleges were surveyed? Why?

**Exercise 8.4**

From a stack of *IEEE Spectrum* magazines, announcements for 84 upcoming engineering conferences were randomly picked. The average length of the conferences was 3.94 days, with a standard deviation of 1.28 days. Assume the underlying population is normal.

- Define the Random Variables  $X$  and  $\bar{X}$ , in words.
- Which distribution should you use for this problem? Explain your choice.
- Construct a 95% confidence interval for the population average length of engineering conferences.
  - State the confidence interval.
  - Sketch the graph.
  - Calculate the error bound.

**Exercise 8.5**

(*Solution on p. 135.*)

Suppose that a committee is studying whether or not there is waste of time in our judicial system. It is interested in the average amount of time individuals waste at the courthouse waiting to be called for service. The committee randomly surveyed 81 people. The sample average was 8 hours with a sample standard deviation of 4 hours.

- i.  $\bar{x} = \underline{\hspace{2cm}}$   
ii.  $s_x = \underline{\hspace{2cm}}$   
iii.  $n = \underline{\hspace{2cm}}$   
iv.  $n - 1 = \underline{\hspace{2cm}}$
- Define the Random Variables  $X$  and  $\bar{X}$ , in words.
- Which distribution should you use for this problem? Explain your choice.
- Construct a 95% confidence interval for the population average time wasted.
  - State the confidence interval.
  - Sketch the graph.
  - Calculate the error bound.
- Explain in a complete sentence what the confidence interval means.

**Exercise 8.6**

Suppose that an accounting firm does a study to determine the time needed to complete one person's tax forms. It randomly surveys 100 people. The sample average is 23.6 hours. There is a known standard deviation of 7.0 hours. The population distribution is assumed to be normal.

- i.  $\bar{x} = \underline{\hspace{2cm}}$   
ii.  $\sigma = \underline{\hspace{2cm}}$   
iii.  $s_x = \underline{\hspace{2cm}}$   
iv.  $n = \underline{\hspace{2cm}}$   
v.  $n - 1 = \underline{\hspace{2cm}}$
- Define the Random Variables  $X$  and  $\bar{X}$ , in words.
- Which distribution should you use for this problem? Explain your choice.
- Construct a 90% confidence interval for the population average time to complete the tax forms.
  - State the confidence interval.
  - Sketch the graph.
  - Calculate the error bound.

- e. If the firm wished to increase its level of confidence and keep the error bound the same by taking another survey, what changes should it make?
- f. If the firm did another survey, kept the error bound the same, and only surveyed 49 people, what would happen to the level of confidence? Why?
- g. Suppose that the firm decided that it needed to be at least 96% confident of the population average length of time to within 1 hour. How would the number of people the firm surveys change? Why?

**Exercise 8.7**

(Solution on p. 135.)

A sample of 16 small bags of the same brand of candies was selected. Assume that the population distribution of bag weights is normal. The weight of each bag was then recorded. The mean weight was 2 ounces with a standard deviation of 0.12 ounces. The population standard deviation is known to be 0.1 ounce.

- a. i.  $\bar{x} = \underline{\hspace{2cm}}$   
ii.  $\sigma = \underline{\hspace{2cm}}$   
iii.  $s_x = \underline{\hspace{2cm}}$   
iv.  $n = \underline{\hspace{2cm}}$   
v.  $n - 1 = \underline{\hspace{2cm}}$
- b. Define the Random Variable  $X$ , in words.
- c. Define the Random Variable  $\bar{X}$ , in words.
- d. Which distribution should you use for this problem? Explain your choice.
- e. Construct a 90% confidence interval for the population average weight of the candies.
  - i. State the confidence interval.
  - ii. Sketch the graph.
  - iii. Calculate the error bound.
- f. Construct a 98% confidence interval for the population average weight of the candies.
  - i. State the confidence interval.
  - ii. Sketch the graph.
  - iii. Calculate the error bound.
- g. In complete sentences, explain why the confidence interval in (f) is larger than the confidence interval in (e).
- h. In complete sentences, give an interpretation of what the interval in (f) means.

**Exercise 8.8**

A pharmaceutical company makes tranquilizers. It is assumed that the distribution for the length of time they last is approximately normal. Researchers in a hospital used the drug on a random sample of 9 patients. The effective period of the tranquilizer for each patient (in hours) was as follows: 2.7; 2.8; 3.0; 2.3; 2.3; 2.2; 2.8; 2.1; and 2.4 .

- a. i.  $\bar{x} = \underline{\hspace{2cm}}$   
ii.  $s_x = \underline{\hspace{2cm}}$   
iii.  $n = \underline{\hspace{2cm}}$   
iv.  $n - 1 = \underline{\hspace{2cm}}$
- b. Define the Random Variable  $X$ , in words.
- c. Define the Random Variable  $\bar{X}$ , in words.
- d. Which distribution should you use for this problem? Explain your choice.
- e. Construct a 95% confidence interval for the population average length of time.
  - i. State the confidence interval.
  - ii. Sketch the graph.

- iii. Calculate the error bound.
- f. What does it mean to be “95% confident” in this problem?

**Exercise 8.9**

(Solution on p. 135.)

Suppose that 14 children were surveyed to determine how long they had to use training wheels. It was revealed that they used them an average of 6 months with a sample standard deviation of 3 months. Assume that the underlying population distribution is normal.

- a. i.  $\bar{x} = \underline{\hspace{2cm}}$
- ii.  $s_x = \underline{\hspace{2cm}}$
- iii.  $n = \underline{\hspace{2cm}}$
- iv.  $n - 1 = \underline{\hspace{2cm}}$
- b. Define the Random Variable  $X$ , in words.
- c. Define the Random Variable  $\bar{X}$ , in words.
- d. Which distribution should you use for this problem? Explain your choice.
- e. Construct a 99% confidence interval for the population average length of time using training wheels.
  - i. State the confidence interval.
  - ii. Sketch the graph.
  - iii. Calculate the error bound.
- f. Why would the error bound change if the confidence level was lowered to 90%?

**Exercise 8.10**

Insurance companies are interested in knowing the population percent of drivers who always buckle up before riding in a car.

- a. When designing a study to determine this population proportion, what is the minimum number you would need to survey to be 95% confident that the population proportion is estimated to within 0.03?
- b. If it was later determined that it was important to be more than 95% confident and a new survey was commissioned, how would that affect the minimum number you would need to survey? Why?

**Exercise 8.11**

(Solution on p. 136.)

Suppose that the insurance companies did do a survey. They randomly surveyed 400 drivers and found that 320 claimed to always buckle up. We are interested in the population proportion of drivers who claim to always buckle up.

- a. i.  $x = \underline{\hspace{2cm}}$
- ii.  $n = \underline{\hspace{2cm}}$
- iii.  $p' = \underline{\hspace{2cm}}$
- b. Define the Random Variables  $X$  and  $P'$ , in words.
- c. Which distribution should you use for this problem? Explain your choice.
- d. Construct a 95% confidence interval for the population proportion that claim to always buckle up.
  - i. State the confidence interval.
  - ii. Sketch the graph.
  - iii. Calculate the error bound.
- e. If this survey were done by telephone, list 3 difficulties the companies might have in obtaining random results.

**Exercise 8.12**

Unoccupied seats on flights cause airlines to lose revenue. Suppose a large airline wants to estimate its average number of unoccupied seats per flight over the past year. To accomplish this, the records of 225 flights are randomly selected and the number of unoccupied seats is noted for each of the sampled flights. The sample mean is 11.6 seats and the sample standard deviation is 4.1 seats.

- a. i.  $\bar{x} = \underline{\hspace{2cm}}$
- ii.  $s_x = \underline{\hspace{2cm}}$
- iii.  $n = \underline{\hspace{2cm}}$
- iv.  $n - 1 = \underline{\hspace{2cm}}$
- b. Define the Random Variables  $X$  and  $\bar{X}$ , in words.
- c. Which distribution should you use for this problem? Explain your choice.
- d. Construct a 92% confidence interval for the population average number of unoccupied seats per flight.
  - i. State the confidence interval.
  - ii. Sketch the graph.
  - iii. Calculate the error bound.

**Exercise 8.13**

(Solution on p. 136.)

According to a recent survey of 1200 people, 61% feel that the president is doing an acceptable job. We are interested in the population proportion of people who feel the president is doing an acceptable job.

- a. Define the Random Variables  $X$  and  $P'$ , in words.
- b. Which distribution should you use for this problem? Explain your choice.
- c. Construct a 90% confidence interval for the population proportion of people who feel the president is doing an acceptable job.
  - i. State the confidence interval.
  - ii. Sketch the graph.
  - iii. Calculate the error bound.

**Exercise 8.14**

A survey of the average amount of cents off that coupons give was done by randomly surveying one coupon per page from the coupon sections of a recent San Jose Mercury News. The following data were collected: 20¢; 75¢; 50¢; 65¢; 30¢; 55¢; 40¢; 40¢; 30¢; 55¢; \$1.50; 40¢; 65¢; 40¢. Assume the underlying distribution is approximately normal.

- a. i.  $\bar{x} = \underline{\hspace{2cm}}$
- ii.  $s_x = \underline{\hspace{2cm}}$
- iii.  $n = \underline{\hspace{2cm}}$
- iv.  $n - 1 = \underline{\hspace{2cm}}$
- b. Define the Random Variables  $X$  and  $\bar{X}$ , in words.
- c. Which distribution should you use for this problem? Explain your choice.
- d. Construct a 95% confidence interval for the population average worth of coupons.
  - i. State the confidence interval.
  - ii. Sketch the graph.
  - iii. Calculate the error bound.
- e. If many random samples were taken of size 14, what percent of the confident intervals constructed should contain the population average worth of coupons? Explain why.

**Exercise 8.15**

(Solution on p. 136.)

An article regarding interracial dating and marriage recently appeared in the *Washington Post*. Of the 1709 randomly selected adults, 315 identified themselves as Latinos, 323 identified themselves as blacks, 254 identified themselves as Asians, and 779 identified themselves as whites. In this survey, 86% of blacks said that their families would welcome a white person into their families. Among Asians, 77% would welcome a white person into their families, 71% would welcome a Latino, and 66% would welcome a black person.

- We are interested in finding the 95% confidence interval for the percent of black families that would welcome a white person into their families. Define the Random Variables  $X$  and  $P'$ , in words.
- Which distribution should you use for this problem? Explain your choice.
- Construct a 95% confidence interval
  - State the confidence interval.
  - Sketch the graph.
  - Calculate the error bound.

**Exercise 8.16**

Refer to the problem above.

- Construct the 95% confidence intervals for the three Asian responses.
- Even though the three point estimates are different, do any of the confidence intervals overlap? Which?
- For any intervals that do overlap, in words, what does this imply about the significance of the differences in the true proportions?
- For any intervals that do not overlap, in words, what does this imply about the significance of the differences in the true proportions?

**Exercise 8.17**

(Solution on p. 136.)

A camp director is interested in the average number of letters each child sends during his/her camp session. The population standard deviation is known to be 2.5. A survey of 20 campers is taken. The average from the sample is 7.9 with a sample standard deviation of 2.8.

- $\bar{x} = \underline{\hspace{2cm}}$
  - $\sigma = \underline{\hspace{2cm}}$
  - $s_x = \underline{\hspace{2cm}}$
  - $n = \underline{\hspace{2cm}}$
  - $n - 1 = \underline{\hspace{2cm}}$
- Define the Random Variables  $X$  and  $\bar{X}$ , in words.
- Which distribution should you use for this problem? Explain your choice.
- Construct a 90% confidence interval for the population average number of letters campers send home.
  - State the confidence interval.
  - Sketch the graph.
  - Calculate the error bound.
- What will happen to the error bound and confidence interval if 500 campers are surveyed? Why?

**Exercise 8.18**

Stanford University conducted a study of whether running is healthy for men and women over age 50. During the first eight years of the study, 1.5% of the 451 members of the 50-Plus Fitness Association died. We are interested in the proportion of people over 50 who ran and died in the same eight-year period.

- a. Define the Random Variables  $X$  and  $P'$ , in words.
- b. Which distribution should you use for this problem? Explain your choice.
- c. Construct a 97% confidence interval for the population proportion of people over 50 who ran and died in the same eight-year period.
  - i. State the confidence interval.
  - ii. Sketch the graph.
  - iii. Calculate the error bound.
- d. Explain what a “97% confidence interval” means for this study.

**Exercise 8.19**

(Solution on p. 136.)

In a recent sample of 84 used cars sales costs, the sample mean was \$6425 with a standard deviation of \$3156. Assume the underlying distribution is approximately normal.

- a. Which distribution should you use for this problem? Explain your choice.
- b. Define the Random Variable  $\bar{X}$ , in words.
- c. Construct a 95% confidence interval for the population average cost of a used car.
  - i. State the confidence interval.
  - ii. Sketch the graph.
  - iii. Calculate the error bound.
- d. Explain what a “95% confidence interval” means for this study.

**Exercise 8.20**

A telephone poll of 1000 adult Americans was reported in an issue of *Time Magazine*. One of the questions asked was “What is the main problem facing the country?” 20% answered “crime”. We are interested in the population proportion of adult Americans who feel that crime is the main problem.

- a. Define the Random Variables  $X$  and  $P'$ , in words.
- b. Which distribution should you use for this problem? Explain your choice.
- c. Construct a 95% confidence interval for the population proportion of adult Americans who feel that crime is the main problem.
  - i. State the confidence interval.
  - ii. Sketch the graph.
  - iii. Calculate the error bound.
- d. Suppose we want to lower the sampling error. What is one way to accomplish that?
- e. The sampling error given by Yankelovich Partners, Inc. (which conducted the poll) is  $\pm 3\%$ . In 1-3 complete sentences, explain what the  $\pm 3\%$  represents.

**Exercise 8.21**

(Solution on p. 136.)

Refer to the above problem. Another question in the poll was “[How much are] you worried about the quality of education in our schools?” 63% responded “a lot”. We are interested in the population proportion of adult Americans who are worried a lot about the quality of education in our schools.

1. Define the Random Variables  $X$  and  $P'$ , in words.
2. Which distribution should you use for this problem? Explain your choice.
3. Construct a 95% confidence interval for the population proportion of adult Americans worried a lot about the quality of education in our schools.
  - i. State the confidence interval.
  - ii. Sketch the graph.

- iii. Calculate the error bound.
4. The sampling error given by Yankelovich Partners, Inc. (which conducted the poll) is  $\pm 3\%$ . In 1-3 complete sentences, explain what the  $\pm 3\%$  represents.

**Exercise 8.22**

Six different national brands of chocolate chip cookies were randomly selected at the supermarket. The grams of fat per serving are as follows: 8; 8; 10; 7; 9; 9. Assume the underlying distribution is approximately normal.

- Calculate a 90% confidence interval for the population average grams of fat per serving of chocolate chip cookies sold in supermarkets.
  - State the confidence interval.
  - Sketch the graph.
  - Calculate the error bound.
- If you wanted a smaller error bound while keeping the same level of confidence, what should have been changed in the study before it was done?
- Go to the store and record the grams of fat per serving of six brands of chocolate chip cookies.
- Calculate the average.
- Is the average within the interval you calculated in part (a)? Did you expect it to be? Why or why not?

**Exercise 8.23**

A confidence interval for a proportion is given to be  $(-0.22, 0.34)$ . Why doesn't the lower limit of the confidence interval make practical sense? How should it be changed? Why?

**8.2.1 Try these multiple choice questions.**

**The next three problems refer to the following:** According a Field Poll conducted February 8 – 17, 2005, 79% of California adults (actual results are 400 out of 506 surveyed) feel that "education and our schools" is one of the top issues facing California. We wish to construct a 90% confidence interval for the true proportion of California adults who feel that education and the schools is one of the top issues facing California.

**Exercise 8.24**

*(Solution on p. 137.)*

A point estimate for the true population proportion is:

- A. 0.90
- B. 1.27
- C. 0.79
- D. 400

**Exercise 8.25**

*(Solution on p. 137.)*

A 90% confidence interval for the population proportion is:

- A.  $(0.761, 0.820)$
- B.  $(0.125, 0.188)$
- C.  $(0.755, 0.826)$
- D.  $(0.130, 0.183)$

**Exercise 8.26**

*(Solution on p. 137.)*

The error bound is approximately

- A. 1.581
- B. 0.791
- C. 0.059
- D. 0.030

**The next two problems refer to the following:**

A quality control specialist for a restaurant chain takes a random sample of size 12 to check the amount of soda served in the 16 oz. serving size. The sample average is 13.30 with a sample standard deviation is 1.55. Assume the underlying population is normally distributed.

**Exercise 8.27**

*(Solution on p. 137.)*

Find the 95% Confidence Interval for the true population mean for the amount of soda served.

- A. (12.42, 14.18)
- B. (12.32, 14.29)
- C. (12.50, 14.10)
- D. Impossible to determine

**Exercise 8.28**

*(Solution on p. 137.)*

What is the error bound?

- A. 0.87
- B. 1.98
- C. 0.99
- D. 1.74

**Exercise 8.29**

*(Solution on p. 137.)*

What is meant by the term “90% confident” when constructing a confidence interval for a mean?

- A. If we took repeated samples, approximately 90% of the samples would produce the same confidence interval.
- B. If we took repeated samples, approximately 90% of the confidence intervals calculated from those samples would contain the sample mean.
- C. If we took repeated samples, approximately 90% of the confidence intervals calculated from those samples would contain the true value of the population mean.
- D. If we took repeated samples, the sample mean would equal the population mean in approximately 90% of the samples.

**The next two problems refer to the following:**

Five hundred and eleven (511) homes in a certain southern California community are randomly surveyed to determine if they meet minimal earthquake preparedness recommendations. One hundred seventy-three (173) of the homes surveyed met the minimum recommendations for earthquake preparedness and 338 did not.

**Exercise 8.30**

*(Solution on p. 137.)*

Find the Confidence Interval at the 90% Confidence Level for the true population proportion of southern California community homes meeting at least the minimum recommendations for earthquake preparedness.

- A. (0.2975, 0.3796)
- B. (0.6270, 6959)
- C. (0.3041, 0.3730)

- D. (0.6204, 0.7025)

**Exercise 8.31**

(*Solution on p. 137.*)

The point estimate for the population proportion of homes that do not meet the minimum recommendations for earthquake preparedness is:

- A. 0.6614
- B. 0.3386
- C. 173
- D. 338

### 8.3 Review Questions<sup>4</sup>

**The next three problems refer to the following situation:** Suppose that a sample of 15 randomly chosen people were put on a special weight loss diet. The amount of weight lost, in pounds, follows an unknown distribution with mean equal to 12 pounds and standard deviation equal to 3 pounds.

**Exercise 8.32: REVIEW QUESTION 1**

(*Solution on p. 137.*)

To find the probability that the average of the 15 people lose no more than 14 pounds, the random variable should be:

- A. The number of people who lost weight on the special weight loss diet
- B. The number of people who were on the diet
- C. The average amount of weight lost by 15 people on the special weight loss diet
- D. The total amount of weight lost by 15 people on the special weight loss diet

**Exercise 8.33: REVIEW QUESTION 2**

(*Solution on p. 137.*)

Find the probability asked for in the previous problem.

**Exercise 8.34: REVIEW QUESTION 3**

(*Solution on p. 137.*)

Find the 90th percentile for the average amount of weight lost by 15 people.

**The next three questions refer to the following situation:** The time of occurrence of the first accident during rush-hour traffic at a major intersection is uniformly distributed between the three hour interval 4 p.m. to 7 p.m. Let  $X$  = the amount of time (hours) it takes for the first accident to occur.

- So, if an accident occurs at 4 p.m., the amount of time, in hours, it took for the accident to occur is \_\_\_\_\_.
- $\mu =$  \_\_\_\_\_
- $\sigma^2 =$  \_\_\_\_\_

**Exercise 8.35: REVIEW QUESTION 4**

(*Solution on p. 137.*)

What is the probability that the time of occurrence is within the first half-hour or the last hour of the period from 4 to 7 p.m.?

- A. Cannot be determined from the information given
- B.  $\frac{1}{6}$
- C.  $\frac{1}{2}$
- D.  $\frac{1}{3}$

**Exercise 8.36: REVIEW QUESTION 5**

(*Solution on p. 137.*)

The 20th percentile occurs after how many hours?

---

<sup>4</sup>This content is available online at <<http://cnx.org/content/m19018/1.1/>>.

- A. 0.20
- B. 0.60
- C. 0.50
- D. 1

**Exercise 8.37: REVIEW QUESTION 6***(Solution on p. 137.)*

Assume Ramon has kept track of the times for the first accidents to occur for 40 different days. Let  $C$  = the total cumulative time. Then  $C$  follows which distribution?

- A.  $U(0, 3)$
- B.  $\text{Exp}\left(\frac{1}{3}\right)$
- C.  $N(60, 30)$
- D.  $N(1.5, 0.01875)$

**Exercise 8.38: REVIEW QUESTION 7***(Solution on p. 137.)*

Using the information in question #6, find the probability that the total time for all first accidents to occur is more than 43 hours.

**The next two questions refer to the following situation:** The length of time a parent must wait for his children to clean their rooms is uniformly distributed in the time interval from 1 to 15 days.

**Exercise 8.39: REVIEW QUESTION 8***(Solution on p. 137.)*

How long must a parent expect to wait for his children to clean their rooms?

- A. 8 days
- B. 3 days
- C. 14 days
- D. 6 days

**Exercise 8.40: REVIEW QUESTION 9***(Solution on p. 137.)*

What is the probability that a parent will wait more than 6 days given that the parent has already waited more than 3 days?

- A. 0.5174
- B. 0.0174
- C. 0.7500
- D. 0.2143

**The next five problems refer to the following study:** Twenty percent of the students at a local community college live in within five miles of the campus. Thirty percent of the students at the same community college receive some kind of financial aid. Of those who live within five miles of the campus, 75% receive some kind of financial aid.

**Exercise 8.41: REVIEW QUESTION 10***(Solution on p. 137.)*

Find the probability that a randomly chosen student at the local community college does not live within five miles of the campus.

- A. 80%
- B. 20%
- C. 30%
- D. Cannot be determined

**Exercise 8.42: REVIEW QUESTION 11***(Solution on p. 137.)*

Find the probability that a randomly chosen student at the local community college lives within five miles of the campus or receives some kind of financial aid.

- A. 50%
- B. 35%
- C. 27.5%
- D. 75%

**Exercise 8.43: REVIEW QUESTION 12***(Solution on p. 137.)*

Based upon the above information, are living in student housing within five miles of the campus and receiving some kind of financial aid mutually exclusive?

- A. Yes
- B. No
- C. Cannot be determined

**Exercise 8.44: REVIEW QUESTION 13***(Solution on p. 137.)*

The interest rate charged on the financial aid is \_\_\_\_\_ data.

- A. quantitative discrete
- B. quantitative continuous
- C. qualitative discrete
- D. qualitative

**Exercise 8.45: REVIEW QUESTION 14***(Solution on p. 137.)*

What follows is information about the students who receive financial aid at the local community college.

- 1st quartile = \$250
- 2nd quartile = \$700
- 3rd quartile = \$1200

(These amounts are for the school year.) If a sample of 200 students is taken, how many are expected to receive \$250 or more?

- A. 50
- B. 250
- C. 150
- D. Cannot be determined

**The next two problems refer to the following information:**  $P(A) = 0.2$ ,  $P(B) = 0.3$ ,  $A$  and  $B$  are independent events.

**Exercise 8.46: REVIEW QUESTION 15***(Solution on p. 137.)*

$P(A \text{ AND } B) =$

- A. 0.5
- B. 0.6
- C. 0
- D. 0.06

**Exercise 8.47: REVIEW QUESTION 16***(Solution on p. 137.)*

$P(A \text{ OR } B) =$

- A. 0.56
- B. 0.5
- C. 0.44

D. 1

**Exercise 8.48: REVIEW QUESTION 17**

*(Solution on p. 137.)*

If  $H$  and  $D$  are mutually exclusive events,  $P(H) = 0.25$ ,  $P(D) = 0.15$ , then  $P(H|D)$

- A. 1
- B. 0
- C. 0.40
- D. 0.0375

## Solutions to Exercises in Chapter 8

### Solution to Exercise 8.1 (p. 121)

- a. i. 71
- ii. 3
- iii. 2.8
- iv. 48
- v.  $\frac{47}{\sqrt{48}}$
- c.  $N\left(71, \frac{3}{\sqrt{48}}\right)$
- d. i. CI: (70.15, 71.85)
- iii. EB = 0.85

### Solution to Exercise 8.3 (p. 122)

- a. i. 8629
- ii. 6944
- iii. 35
- iv. 34
- c.  $t_{34}$
- d. i. CI: (6243, 11,014)
- iii. EB = 2385
- e. It will become smaller

### Solution to Exercise 8.5 (p. 123)

- a. i. 8
- ii. 4
- iii. 81
- iv. 80
- c.  $t_{80}$
- d. i. CI: (7.12, 8.88)
- iii. EB = 0.88

### Solution to Exercise 8.7 (p. 124)

- a. i. 2
- ii. 0.1
- iii. 0.12
- iv. 16
- v. 15
- b. the weight of 1 small bag of candies
- c. the average weight of 16 small bags of candies
- d.  $N\left(2, \frac{0.1}{\sqrt{16}}\right)$
- e. i. CI: (1.96, 2.04)
- iii. EB = 0.04
- f. i. CI: (1.94, 2.06)
- iii. EB = 0.06

### Solution to Exercise 8.9 (p. 125)

- a. i. 6
- ii. 3
- iii. 14

- iv. 13  
 b. the time for a child to remove his training wheels  
 c. the average time for 14 children to remove their training wheels.  
 d.  $t_{13}$   
 e. i. CI: (3.58, 8.42)  
     iii. EB = 2.42

**Solution to Exercise 8.11 (p. 125)**

- a. i. 320  
     ii. 400  
     iii. 0.80  
 c.  $N\left(0.80, \sqrt{\frac{(0.80)(0.20)}{400}}\right)$   
 d. i. CI: (0.76, 0.84)  
     iii. EB = 0.04

**Solution to Exercise 8.13 (p. 126)**

- b.  $N\left(0.61, \sqrt{\frac{(0.61)(0.39)}{1200}}\right)$   
 c. i. CI: (0.59, 0.63)  
     iii. EB = 0.02

**Solution to Exercise 8.15 (p. 127)**

- b.  $N\left(0.86, \sqrt{\frac{(0.86)(0.14)}{323}}\right)$   
 c. i. CI: (0.8229, 0.8984)  
     iii. EB = 0.038

**Solution to Exercise 8.17 (p. 127)**

- a. i. 7.9  
     ii. 2.5  
     iii. 2.8  
     iv. 20  
     v. 19  
 c.  $N\left(7.9, \frac{2.5}{\sqrt{20}}\right)$   
 d. i. CI: (6.98, 8.82)  
     iii. EB: 0.92

**Solution to Exercise 8.19 (p. 128)**

- a.  $t_{83}$   
 b. average cost of 84 used cars  
 c. i. CI: (5740.10, 7109.90)  
     iii. EB = 684.90

**Solution to Exercise 8.21 (p. 128)**

- b.  $N\left(0.63, \sqrt{\frac{(0.63)(0.37)}{1000}}\right)$   
 c. i. CI: (0.60, 0.66)  
     iii. EB = 0.03

**Solution to Exercise 8.24 (p. 129)**

C

**Solution to Exercise 8.25 (p. 129)**

A

**Solution to Exercise 8.26 (p. 129)**

D

**Solution to Exercise 8.27 (p. 130)**

B

**Solution to Exercise 8.28 (p. 130)**

C

**Solution to Exercise 8.29 (p. 130)**

C

**Solution to Exercise 8.30 (p. 130)**

C

**Solution to Exercise 8.31 (p. 131)**

A

**Solution to Exercise 8.32 (p. 131)**

REVIEW QUESTION 1 Solution : C

**Solution to Exercise 8.33 (p. 131)**

REVIEW QUESTION 2 Solution : 0.9951

**Solution to Exercise 8.34 (p. 131)**

REVIEW QUESTION 3 Solution : 12.99

**Solution to Exercise 8.35 (p. 131)**

REVIEW QUESTION 4 Solution : C

**Solution to Exercise 8.36 (p. 131)**

REVIEW QUESTION 5 Solution : B

**Solution to Exercise 8.37 (p. 132)**

REVIEW QUESTION 6 Solution : C

**Solution to Exercise 8.38 (p. 132)**

REVIEW QUESTION 7 Solution : 0.9990

**Solution to Exercise 8.39 (p. 132)**

REVIEW QUESTION 8 Solution : A

**Solution to Exercise 8.40 (p. 132)**

REVIEW QUESTION 9 Solution : C

**Solution to Exercise 8.41 (p. 132)**

REVIEW QUESTION 10 Solution : A

**Solution to Exercise 8.42 (p. 132)**

REVIEW QUESTION 11 Solution : B

**Solution to Exercise 8.43 (p. 133)**

REVIEW QUESTION 12 Solution : B

**Solution to Exercise 8.44 (p. 133)**

REVIEW QUESTION 13 Solution : B

**Solution to Exercise 8.45 (p. 133)**

REVIEW QUESTION 14 Solution : C. 150

**Solution to Exercise 8.46 (p. 133)**

REVIEW QUESTION 15 Solution : D

**Solution to Exercise 8.47 (p. 133)**

REVIEW QUESTION 16 Solution : C

**Solution to Exercise 8.48 (p. 134)**

REVIEW QUESTION 17 Solution : B



## Chapter 9

# Hypothesis Testing: Single Mean and Single Proportion

### 9.1 Summary of Formulas<sup>1</sup>

$H_o$  and  $H_a$  are contradictory.

If $H_o$ has:	equal ( $=$ )	greater than or equal to ( $\geq$ )	less than or equal to ( $\leq$ )
then $H_a$ has:	not equal ( $\neq$ ) or greater than ( $>$ ) or less than ( $<$ )	less than ( $<$ )	greater than ( $>$ )

Table 9.1

If  $\alpha \leq$  p-value, then do not reject  $H_o$ .

If  $\alpha >$  p-value, then reject  $H_o$ .

$\alpha$  is preconceived. Its value is set before the hypothesis test starts. The p-value is calculated from the data.

$\alpha$  = probability of a Type I error =  $P(\text{Type I error})$  = probability of rejecting the null hypothesis when the null hypothesis is true.

$\beta$  = probability of a Type II error =  $P(\text{Type II error})$  = probability of not rejecting the null hypothesis when the null hypothesis is false.

If there is no given preconceived  $\alpha$ , then use  $\alpha = 0.05$ .

#### Types of Hypothesis Tests

- Single population mean, **known** population variance (or standard deviation): **Normal test**.
- Single population mean, **unknown** population variance (or standard deviation): **Student-t test**.
- Single population proportion: **Normal test**.

<sup>1</sup>This content is available online at <<http://cnx.org/content/m16996/1.7/>>.

## 9.2 Homework (modified R. Bloom)<sup>2</sup>

Some exercises from the original version of this textbook have been removed in this revision of this section: #11,12,14,18-24,26,27. They are available at <http://cnx.org/content/m17001/>

### Exercise 9.1

(*Solution on p. 149.*)

Some of the statements below refer to the null hypothesis, some to the alternate hypothesis.

State the null hypothesis,  $H_0$ , and the alternative hypothesis,  $H_a$ , in terms of the appropriate parameter ( $\mu$  or  $p$ ).

- a. Americans work an average of 34 years before retiring.
- b. At most 60% of Americans vote in presidential elections.
- c. The average starting salary for San Jose State University graduates is at least \$100,000 per year.
- d. 29% of high school seniors get drunk each month.
- e. Fewer than 5% of adults ride the bus to work in Los Angeles.
- f. The average number of cars a person owns in her lifetime is not more than 10.
- g. About half of Americans prefer to live away from cities, given the choice.
- h. Europeans have an average paid vacation each year of six weeks.
- i. The chance of developing breast cancer is under 11% for women.
- j. Private universities cost, on average, more than \$20,000 per year for tuition.

### Exercise 9.2

(*Solution on p. 149.*)

For (a) - (j) above, state the Type I and Type II errors in complete sentences.

### Exercise 9.3

For (a) - (j) above, in complete sentences:

- a. State a consequence of committing a Type I error.
- b. State a consequence of committing a Type II error.

NOTE: For each of the word problems, use a solution sheet to do the hypothesis test.

NOTE: If you are using a student-t distribution for a homework problem below, you may assume that the underlying population is normally distributed. (However, in general, a statistician would first need to verify that this assumption is reasonable before applying a t-test.)

### Exercise 9.4

A particular brand of tires claims that its deluxe tire averages at least 50,000 miles before it needs to be replaced. From past studies of this tire, the standard deviation is known to be 8000. A survey of owners of that tire design is conducted. From the 28 tires surveyed, the average lifespan was 46,500 miles with a standard deviation of 9800 miles. Do the data support the claim at the 5% level?

### Exercise 9.5

(*Solution on p. 149.*)

From generation to generation, the average age when smokers first start to smoke varies. However, the standard deviation of that age remains constant of around 2.1 years. A survey of 40 smokers of this generation was done to see if the average starting age is at least 19. The sample average was 18.1 with a sample standard deviation of 1.3. Do the data support the claim at the 5% level?

---

<sup>2</sup>This content is available online at <<http://cnx.org/content/m18867/1.1/>>.

**Exercise 9.6**

The cost of a daily newspaper varies from city to city. However, the variation among prices remains steady with a standard deviation of 6¢. A study was done to test the claim that the average cost of a daily newspaper is 35¢. Twelve costs yield an average cost of 30¢ with a standard deviation of 4¢. Do the data support the claim at the 1% level?

**Exercise 9.7**

*(Solution on p. 149.)*

An article in the **San Jose Mercury News** stated that students in the California state university system take an average of 4.5 years to finish their undergraduate degrees. Suppose you believe that the average time is longer. You conduct a survey of 49 students and obtain a sample mean of 5.1 with a sample standard deviation of 1.2. Do the data support your claim at the 1% level?

**Exercise 9.8**

The average number of sick days an employee takes per year is believed to be about 10. Members of a personnel department do not believe this figure. They randomly survey 8 employees. The number of sick days they took for the past year are as follows: 12; 4; 15; 3; 11; 8; 6; 8. Let  $x$  = the number of sick days they took for the past year. Should the personnel team believe that the average number is about 10?

**Exercise 9.9**

*(Solution on p. 149.)*

In 1955, **Life Magazine** reported that the 25 year-old mother of three worked [on average] an 80 hour week. Recently, many groups have been studying whether or not the women's movement has, in fact, resulted in an increase in the average work week for women (combining employment and at-home work). Suppose a study was done to determine if the average work week has increased. 81 women were surveyed with the following results. The sample average was 83 hours; the sample standard deviation was 10 hours. Does it appear that the average work week has increased for women at the 5% level?

**Exercise 9.10**

Your statistics instructor claims that 60 percent of the students who take her Elementary Statistics class go through life feeling more enriched. For some reason that she can't quite figure out, most people don't believe her. You decide to check this out on your own. You randomly survey 64 of her past Elementary Statistics students and find that 34 feel more enriched as a result of her class. Now, what do you think?

**Exercise 9.11**

Exercise 11 removed from textbook

**Exercise 9.12**

Exercise 12 removed from textbook

**Exercise 9.13**

*(Solution on p. 149.)*

According to an article in **Newsweek**, the natural ratio of girls to boys is 100:105. In China, the birth ratio is 100: 114 (46.7% girls). Suppose you don't believe the reported figures of the percent of girls born in China. You conduct a study. In this study, you count the number of girls and boys born in 150 randomly chosen recent births. There are 60 girls and 90 boys born of the 150. Based on your study, do you believe that the percent of girls born in China is 46.7?

**Exercise 9.14**

Exercise 14 removed from textbook

**Exercise 9.15**

*(Solution on p. 149.)*

The average work week for engineers in a start-up company is believed to be about 60 hours. A newly hired engineer hopes that it's shorter. She asks a random sample of 10 engineers working in start-ups for the lengths of their average work weeks. Based on the results that follow, should she count on the average work week to be shorter than 60 hours?

Data (length of average work week): 70; 45; 55; 60; 65; 55; 55; 60; 50; 55.

**Exercise 9.16**

Use the “Lap time” data for Lap 4 (see Table of Contents) to test the claim that Terri finishes Lap 4 on average in less than 129 seconds. Use all twenty races given.

**Exercise 9.17**

Use the “Initial Public Offering” data (see Table of Contents) to test the claim that the average offer price was \$18 per share. Do not use all the data. Use your random number generator to randomly survey 15 prices.

**Exercise 9.18**

Exercise removed from textbook.

**Exercise 9.19**

Exercise removed from textbook

**Exercise 9.20**

Exercise removed from textbook

**Exercise 9.21**

Exercise removed from textbook

**Exercise 9.22**

Exercise removed from textbook

**Exercise 9.23**

Exercise removed from textbook

**Exercise 9.24**

Exercise removed from textbook

**Exercise 9.25**

Japanese Girls’ Names, by Kumi Furuichi

(*Solution on p. 150.*)

It used to be very typical for Japanese girls’ names to end with “ko.” (The trend might have started around my grandmothers’ generation and its peak might have been around my mother’s generation.) “Ko” means “child” in Chinese character. Parents would name their daughters with “ko” attaching to other Chinese characters which have meanings that they want their daughters to become, such as Sachiko – a happy child, Yoshiko – a good child, Yasuko – a healthy child, and so on.

However, I noticed recently that only two out of nine of my Japanese girlfriends at this school have names which end with “ko.” More and more, parents seem to have become creative, modernized, and, sometimes, westernized in naming their children.

I have a feeling that, while 70 percent or more of my mother’s generation would have names with “ko” at the end, the proportion has dropped among my peers. I wrote down all my Japanese friends’, ex-classmates’, co-workers, and acquaintances’ names that I could remember. Below are the names. (Some are repeats.) Test to see if the proportion has dropped for this generation.

Ai, Akemi, Akiko, Ayumi, Chiaki, Chie, Eiko, Eri, Eriko, Fumiko, Harumi, Hitomi, Hiroko, Hiroko, Hidemi, Hisako, Hinako, Izumi, Izumi, Junko, Junko, Kana, Kanako, Kanayo, Kayo, Kayoko, Kazumi, Keiko, Keiko, Kei, Kumi, Kumiko, Kyoko, Kyoko, Madoka, Maho, Mai, Maiko, Maki, Miki, Miki, Mikiko, Mina, Minako, Miyako, Momoko, Nana, Naoko, Naoko, Naoko, Noriko, Rieko, Rika, Rika, Rumiko, Rei, Reiko, Reiko, Sachiko, Sachiko, Sachiyo, Saki, Sayaka, Sayoko, Sayuri, Seiko, Shiho, Shizuka, Sumiko, Takako, Takako, Tomoe, Tomoe, Tomoko, Touko, Yasuko, Yasuko, Yasuyo, Yoko, Yoko, Yoko, Yoshiko, Yoshiko, Yoshiko, Yuka, Yuki, Yuki, Yukiko, Yuko, Yuko.

**Exercise 9.26**

Exercise removed from textbook

**Exercise 9.27**

Exercise removed from textbook

**Exercise 9.28**

Toastmasters International cites a February 2001 report by Gallop Poll that 40% of Americans fear public speaking. A student believes that less than 40% of students at her school fear public speaking. She randomly surveys 361 schoolmates and finds that 135 report they fear public speaking. Conduct a hypothesis test to determine if the percent at her school is less than 40%. (Source: <http://toastmasters.org/artisan/detail.asp?CategoryID=1&SubCategoryID=10&ArticleID=429&Page=1><sup>3</sup>)

**Exercise 9.29**

(*Solution on p. 150.*)

In 2004, 68% of online courses taught at community colleges nationwide were taught by full-time faculty. To test if 68% also represents California's percent for full-time faculty teaching the online classes, Long Beach City College (LBCC), CA, was randomly selected for comparison. In 2004, 34 of the 44 online courses LBCC offered were taught by full-time faculty. Conduct a hypothesis test to determine if 68% represents CA. NOTE: For a true test, use more CA community colleges. (Sources: **Growing by Degrees** by Allen and Seaman; Amit Schitai, Director of Instructional Technology and Distance Learning, LBCC).

NOTE: For a true test, use more CA community colleges.

**Exercise 9.30**

According to an article in **The New York Times** (5/12/2004), 19.3% of New York City adults smoked in 2003. Suppose that a survey is conducted to determine this year's rate. Twelve out of 70 randomly chosen N.Y. City residents reply that they smoke. Conduct a hypothesis test to determine if the rate is still 19.3%.

**Exercise 9.31**

(*Solution on p. 150.*)

The average age of De Anza College students in Winter 2006 term was 26.6 years old. An instructor thinks the average age for online students is older than 26.6. She randomly surveys 56 online students and finds that the sample average is 29.4 with a standard deviation of 2.1. Conduct a hypothesis test. (Source: [http://research.fhda.edu/factbook/DAdemofs/Fact\\_sheet\\_da\\_2006w.pdf](http://research.fhda.edu/factbook/DAdemofs/Fact_sheet_da_2006w.pdf)<sup>4</sup>)

**Exercise 9.32**

In 2004, registered nurses earned an average annual salary of \$52,330. A survey was conducted of 41 California nurses to determine if the annual salary is higher than \$52,330 for California nurses. The sample average was \$61,121 with a sample standard deviation of \$7,489. Conduct a hypothesis test. (Source: <http://stats.bls.gov/oco/ocos083.htm#earnings><sup>5</sup>)

**Exercise 9.33**

(*Solution on p. 150.*)

La Leche League International reports that the average age of weaning a child from breastfeeding is age 4 to 5 worldwide. In America, most nursing mothers wean their children much earlier. Suppose a random survey is conducted of 21 U.S. mothers who recently weaned their children. The average weaning age was 9 months ( $\frac{3}{4}$  year) with a standard deviation of 4 months. Conduct a hypothesis test to determine if the average weaning age in the U.S. is less than 4 years old. (Source: <http://www.lalecheleague.org/Law/BAFeb01.html><sup>6</sup>)

<sup>3</sup><http://toastmasters.org/artisan/detail.asp?CategoryID=1&SubCategoryID=10&ArticleID=429&Page=1>

<sup>4</sup>[http://research.fhda.edu/factbook/DAdemofs/Fact\\_sheet\\_da\\_2006w.pdf](http://research.fhda.edu/factbook/DAdemofs/Fact_sheet_da_2006w.pdf)

<sup>5</sup><http://stats.bls.gov/oco/ocos083.htm#earnings>

<sup>6</sup><http://www.lalecheleague.org/Law/BAFeb01.html>

### 9.2.1 Try these multiple choice questions.

**Exercise 9.34**
*(Solution on p. 150.)*

When a new drug is created, the pharmaceutical company must subject it to testing before receiving the necessary permission from the Food and Drug Administration (FDA) to market the drug. Suppose the null hypothesis is “the drug is unsafe.” What is the Type II Error?

- A. To claim the drug is safe when in, fact, it is unsafe
- B. To claim the drug is unsafe when, in fact, it is safe.
- C. To claim the drug is safe when, in fact, it is safe.
- D. To claim the drug is unsafe when, in fact, it is unsafe

**The next two questions refer to the following information:** Over the past few decades, public health officials have examined the link between weight concerns and teen girls smoking. Researchers surveyed a group of 273 randomly selected teen girls living in Massachusetts (between 12 and 15 years old). After four years the girls were surveyed again. Sixty-three (63) said they smoked to stay thin. Is there good evidence that more than thirty percent of the teen girls smoke to stay thin?

**Exercise 9.35**
*(Solution on p. 150.)*

The alternate hypothesis is

- A.  $p < 0.30$
- B.  $p \leq 0.30$
- C.  $p \geq 0.30$
- D.  $p > 0.30$

**Exercise 9.36**
*(Solution on p. 150.)*

After conducting the test, your decision and conclusion are

- A. Reject  $H_0$ : More than 30% of teen girls smoke to stay thin.
- B. Do not reject  $H_0$ : Less than 30% of teen girls smoke to stay thin.
- C. Do not reject  $H_0$ : At most 30% of teen girls smoke to stay thin.
- D. Reject  $H_0$ : Less than 30% of teen girls smoke to stay thin.

**The next three questions refer to the following information:** A statistics instructor believes that fewer than 20% of Evergreen Valley College (EVC) students attended the opening night midnight showing of the latest Harry Potter movie. For a random sample of 84 EVC students, 11 of the students in the sample attended the midnight showing.

**Exercise 9.37**
*(Solution on p. 150.)*

An appropriate alternative hypothesis is

- A.  $p = 0.20$
- B.  $p > 0.20$
- C.  $p < 0.20$
- D.  $p \leq 0.20$

**Exercise 9.38**
*(Solution on p. 150.)*

At a 1% level of significance, an appropriate conclusion is:

- A. The percent of EVC students who attended the midnight showing of Harry Potter is at least 20%.
- B. The percent of EVC students who attended the midnight showing of Harry Potter is more than 20%.

- C. The percent of EVC students who attended the midnight showing of Harry Potter is less than 20%.
- D. There is not enough information to make a decision.

**Exercise 9.39**

(Solution on p. 150.)

The Type I error is believing that the percent of EVC students who attended is:

- A. at least 20%, when in fact, it is less than 20%.
- B. 20%, when in fact, it is 20%.
- C. less than 20%, when in fact, it is at least 20%.
- D. less than 20%, when in fact, it is less than 20%.

**The next two questions refer to the following information:**

It is believed that Lake Tahoe Community College (LTCC) Intermediate Algebra students get less than 7 hours of sleep per night, on average. A survey of 22 LTCC Intermediate Algebra students generated an average of 7.24 hours with a standard deviation of 1.93 hours. At a level of significance of 5%, do LTCC Intermediate Algebra students get less than 7 hours of sleep per night, on average?

**Exercise 9.40**

(Solution on p. 150.)

The distribution to be used for this test is  $\bar{X} \sim$

- A.  $N\left(7.24, \frac{1.93}{\sqrt{22}}\right)$
- B.  $N(7.24, 1.93)$
- C.  $t_{22}$
- D.  $t_{21}$

**Exercise 9.41**

(Solution on p. 150.)

The Type II error is “I believe that the average number of hours of sleep LTCC students get per night

- A. is less than 7 hours when, in fact, it is at least 7 hours.”
- B. is less than 7 hours when, in fact, it is less than 7 hours.”
- C. is at least 7 hours when, in fact, it is at least 7 hours.”
- D. is at least 7 hours when, in fact, it is less than 7 hours.”

**The next three questions refer to the following information:** An organization in 1995 reported that teenagers spent an average of 4.5 hours per week on the telephone. The organization thinks that, in 2007, the average is higher. Fifteen (15) randomly chosen teenagers were asked how many hours per week they spend on the telephone. The sample mean was 4.75 hours with a sample standard deviation of 2.0.

**Exercise 9.42**

(Solution on p. 150.)

The null and alternate hypotheses are:

- A.  $H_0 : \bar{x} = 4.5, H_a : \bar{x} > 4.5$
- B.  $H_0 : \mu \geq 4.5 H_a : \mu < 4.5$
- C.  $H_0 : \mu = 4.75 H_a : \mu > 4.75$
- D.  $H_0 : \mu = 4.5 H_a : \mu > 4.5$

**Exercise 9.43**

(Solution on p. 150.)

At a significance level of  $\alpha = 0.05$ , the correct conclusion is:

- A. The average in 2007 is higher than it was in 1995.
- B. The average in 1995 is higher than in 2007.

- C. The average is still about the same as it was in 1995.
- D. The test is inconclusive.

**Exercise 9.44**

(Solution on p. 150.)

The Type I error is:

- A. To conclude the average hours per week in 2007 is higher than in 1995, when in fact, it is higher.
- B. To conclude the average hours per week in 2007 is higher than in 1995, when in fact, it is the same.
- C. To conclude the average hours per week in 2007 is the same as in 1995, when in fact, it is higher.
- D. To conclude the average hours per week in 2007 is no higher than in 1995, when in fact, it is not higher.

### 9.3 Review Questions<sup>7</sup>

**Exercise 9.45: REVIEW QUESTION 1**

(Solution on p. 150.)

1. Rebecca and Matt are 14 year old twins. Matt's height is 2 standard deviations below the mean for 14 year old boys' height. Rebecca's height is 0.10 standard deviations above the mean for 14 year old girls' height. Interpret this.

- A. Matt is 2.1 inches shorter than Rebecca
- B. Rebecca is very tall compared to other 14 year old girls.
- C. Rebecca is taller than Matt.
- D. Matt is shorter than the average 14 year old boy.

**Exercise 9.46: REVIEW QUESTION 2**

2. Construct a histogram of the IPO data (see Appendix). Use 5 intervals.

**The next six questions refer to the following information:** Ninety homeowners were asked the number of estimates they obtained before having their homes fumigated. X = the number of estimates.

X	Rel. Freq.	Cumulative Rel. Freq.
1	0.3	
2	0.2	
4	0.4	
5	0.1	

Table 9.2

**Exercise 9.47: REVIEW QUESTION 3**

(Solution on p. 151.)

Calculate the frequencies

**Exercise 9.48: REVIEW QUESTION 4**

(Solution on p. 151.)

Complete the cumulative relative frequency column. What percent of the estimates fell at or below 4?

<sup>7</sup>This content is available online at <<http://cnx.org/content/m19025/1.1/>>.

**Exercise 9.49: REVIEW QUESTION 5**

(Solution on p. 151.)

5. Calculate the sample mean (a) and sample standard deviation (b).

**Exercise 9.50: REVIEW QUESTION 6**

(Solution on p. 151.)

6. Calculate the median, M, the first quartile, Q1, the third quartile, Q3.

**Exercise 9.51: REVIEW QUESTION 7**

(Solution on p. 151.)

7. The middle 50% of the data are between \_\_\_\_\_ and \_\_\_\_\_.

**Exercise 9.52: REVIEW QUESTION 8**

8. Construct a boxplot of the data.

**The next three questions refer to the following table:** Seventy 5th and 6th graders were asked their favorite dinner.

	Pizza	Hamburgers	Spaghetti	Fried shrimp
5th grader	15	6	9	0
6th grader	15	7	10	8

**Table 9.3****Exercise 9.53: REVIEW QUESTION 9**

(Solution on p. 151.)

9. Find the probability that one randomly chosen child is in the 6th grade and prefers fried shrimp.

- A.  $\frac{32}{70}$
- B.  $\frac{8}{32}$
- C.  $\frac{8}{8}$
- D.  $\frac{8}{70}$

**Exercise 9.54: REVIEW QUESTION 10**

(Solution on p. 151.)

10. Find the probability that a child does not prefer pizza.

- A.  $\frac{30}{70}$
- B.  $\frac{30}{40}$
- C.  $\frac{40}{70}$
- D. 1

**Exercise 9.55: REVIEW QUESTION 11**

(Solution on p. 151.)

11. Find the probability a child is in the 5th grade given that the child prefers spaghetti.

- A.  $\frac{9}{19}$
- B.  $\frac{9}{70}$
- C.  $\frac{9}{30}$
- D.  $\frac{19}{70}$

**Exercise 9.56: REVIEW QUESTION 12**

(Solution on p. 151.)

12. A sample of convenience is a random sample.

- A. true
- B. false

**Exercise 9.57: REVIEW QUESTION 13**

(Solution on p. 151.)

13. A statistic is a number that is a property of the population.

- A. true
- B. false

**Exercise 9.58: REVIEW QUESTION 14***(Solution on p. 151.)*

14. You should always throw out any data that are outliers.

- A. true
- B. false

**Exercise 9.59: REVIEW QUESTION 15***(Solution on p. 151.)*

15. Lee bakes pies for a little restaurant in Felton. She generally bakes 20 pies in a day, on the average.

- a. Define the Random Variable X.
- b. State the distribution for X.
- c. Find the probability that Lee bakes more than 25 pies in any given day.

**Exercise 9.60: REVIEW QUESTION 16***(Solution on p. 151.)*

16. Six different brands of Italian salad dressing were randomly selected at a supermarket. The grams of fat per serving are 7, 7, 9, 6, 8, 5. Assume that the underlying distribution is normal. Calculate a 95% confidence interval for the population average grams of fat per serving of Italian salad dressing sold in supermarkets.

**Exercise 9.61: REVIEW QUESTION 17***(Solution on p. 151.)*

17. Given: uniform, exponential, normal distributions. Match each to a statement below.

- a. mean = median  $\neq$  mode
- b. mean > median > mode
- c. mean = median = mode

## Solutions to Exercises in Chapter 9

### Solution to Exercise 9.1 (p. 140)

Complete solutions to all parts of this problem are available on the instructor's website for this class.

- a.  $H_0 : \mu = 34 ; H_a : \mu \neq 34$
- c.  $H_0 : \mu \geq 100,000 ; H_a : \mu < 100,000$
- d.  $H_0 : p = 0.29 ; H_a : p \neq 0.29$
- g.  $H_0 : p = 0.50 ; H_a : p \neq 0.50$
- i.  $H_0 : p \geq 0.11 ; H_a : p < 0.11$

### Solution to Exercise 9.2 (p. 140)

Complete solutions to all parts of this problem are available on the instructor's website for this class.

- a. Type I error: We believe the average is not 34 years, when it really is 34 years. Type II error: We believe the average is 34 years, when it is not really 34 years.
- c. Type I error: We believe the average is less than \$100,000, when it really is at least \$100,000. Type II error: We believe the average is at least \$100,000, when it is really less than \$100,000.
- d. Type I error: We believe that the proportion of h.s. seniors who get drunk each month is not 29%, when it really is 29%. Type II error: We believe that 29% of h.s. seniors get drunk each month, when the proportion is really not 29%.
- i. Type I error: We believe the proportion is less than 11%, when it is really at least 11%. Type II error: We believe the proportion is at least 11%, when it really is less than 11%.

### Solution to Exercise 9.5 (p. 140)

- e.  $z = -2.71$
- f. 0.0034
- h. Decision: Reject null; Conclusion:  $\mu < 19$
- i. (17.449, 18.757)

### Solution to Exercise 9.7 (p. 141)

- e. 3.5
- f. 0.0005
- h. Decision: Reject null; Conclusion:  $\mu > 4.5$
- i. (4.7553, 5.4447)

### Solution to Exercise 9.9 (p. 141)

- e. 2.7
- f. 0.0042
- h. Decision: Reject Null
- i. (80.789, 85.211)

### Solution to Exercise 9.13 (p. 141)

- e. -1.64
- f. 0.1000
- h. Decision: Do not reject null
- i. (0.3216, 0.4784)

### Solution to Exercise 9.15 (p. 141)

- d.  $t_9$
- e. -1.33

- f. 0.1086
- h. Decision: Do not reject null
- i. (51.886, 62.114)

**Solution to Exercise 9.25 (p. 142)**

- e.  $z = -2.99$
- f. 0.0014
- h. Decision: Reject null; Conclusion:  $p < .70$
- i. (0.4529, 0.6582)

**Solution to Exercise 9.29 (p. 143)**

- e. 1.32
- f. 0.1873
- h. Decision: Do not reject null
- i. (0.65, 0.90)

**Solution to Exercise 9.31 (p. 143)**

- e. 9.98
- f. 0.0000
- h. Decision: Reject null
- i. (28.8, 30.0)

**Solution to Exercise 9.33 (p. 143)**

- e. -44.7
- f. 0.0000
- h. Decision: Reject null
- i. (0.60, 0.90) - in years

**Solution to Exercise 9.34 (p. 144)**

B

**Solution to Exercise 9.35 (p. 144)**

D

**Solution to Exercise 9.36 (p. 144)**

C

**Solution to Exercise 9.37 (p. 144)**

C

**Solution to Exercise 9.38 (p. 144)**

A

**Solution to Exercise 9.39 (p. 145)**

C

**Solution to Exercise 9.40 (p. 145)**

D

**Solution to Exercise 9.41 (p. 145)**

D

**Solution to Exercise 9.42 (p. 145)**

D

**Solution to Exercise 9.43 (p. 145)**

C

**Solution to Exercise 9.44 (p. 146)**

B

**Solution to Exercise 9.45 (p. 146)**

REVIEW QUESTION 1 Solution : D

**Solution to Exercise 9.47 (p. 146)**

REVIEW QUESTION 3 Solution

X	Frequency
1	27
2	18
4	36
5	9

**Table 9.4****Solution to Exercise 9.48 (p. 146)**

REVIEW QUESTION 4 Solution : 90%

**Solution to Exercise 9.49 (p. 147)****REVIEW QUESTION 5 Solution**

- a. 2.8
- b. 1.48

**Solution to Exercise 9.50 (p. 147)**REVIEW QUESTION 6 Solution :  $M = 3 ; Q1 = 1 ; Q3 = 4$ **Solution to Exercise 9.51 (p. 147)**

REVIEW QUESTION 7 Solution : 1 and 4

**Solution to Exercise 9.53 (p. 147)**

REVIEW QUESTION 9 Solution : D

**Solution to Exercise 9.54 (p. 147)**

REVIEW QUESTION 10 Solution : C

**Solution to Exercise 9.55 (p. 147)**

REVIEW QUESTION 11 Solution : A

**Solution to Exercise 9.56 (p. 147)**

REVIEW QUESTION 12 Solution : B

**Solution to Exercise 9.57 (p. 147)**

REVIEW QUESTION 13 Solution : B

**Solution to Exercise 9.58 (p. 148)**

REVIEW QUESTION 14 Solution : B

**Solution to Exercise 9.59 (p. 148)****REVIEW QUESTION 15 Solution**

- b.  $P(20)$
- c. 0.1122

**Solution to Exercise 9.60 (p. 148)**

REVIEW QUESTION 16 Solution : CI: (5.52, 8.48)

**Solution to Exercise 9.61 (p. 148)****REVIEW QUESTION 17 Solution**

- a. uniform
- b. exponential
- c. normal



# Chapter 10

## Hypothesis Testing: Two Means, Paired Data, Two Proportions

### 10.1 Summary of Types of Hypothesis Tests<sup>1</sup>

#### Two Population Means

- Populations are independent and population standard deviations are unknown.
- Populations are independent and population standard deviations are known (not likely).

#### Matched or Paired Samples

- Two samples are drawn from the same set of objects.
- Samples are dependent.

#### Two Population Proportions

- Populations are independent.

### 10.2 Homework<sup>2</sup>

For questions Exercise 10.1 - Exercise 10.10, indicate which of the following choices best identifies the hypothesis test.

- A. Independent group means, population standard deviations and/or variances known
- B. Independent group means, population standard deviations and/or variances unknown
- C. Matched or paired samples
- D. Single mean
- E. 2 proportions
- F. Single proportion

#### Exercise 10.1

(*Solution on p. 166.*)

A powder diet is tested on 49 people and a liquid diet is tested on 36 different people. The population standard deviations are 2 pounds and 3 pounds, respectively. Of interest is whether the liquid diet yields a higher average weight loss than the powder diet.

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m17044/1.5/>>.

<sup>2</sup>This content is available online at <<http://cnx.org/content/m17023/1.15/>>.

**Exercise 10.2**

A new chocolate bar is taste-tested on consumers. Of interest is whether the percentage of children that like the new chocolate bar is greater than the percentage of adults that like it.

**Exercise 10.3**

(*Solution on p. 166.*)

The average number of English courses taken in a two-year time period by male and female college students is believed to be about the same. An experiment is conducted and data are collected from 9 males and 16 females.

**Exercise 10.4**

A football league reported that the average number of touchdowns per game was 5. A study is done to determine if the average number of touchdowns has decreased.

**Exercise 10.5**

(*Solution on p. 166.*)

A study is done to determine if students in the California state university system take longer to graduate than students enrolled in private universities. 100 students from both the California state university system and private universities are surveyed. From years of research, it is known that the population standard deviations are 1.5811 years and 1 year, respectively.

**Exercise 10.6**

According to a YWCA Rape Crisis Center newsletter, 75% of rape victims know their attackers. A study is done to verify this.

**Exercise 10.7**

(*Solution on p. 166.*)

According to a recent study, U.S. companies have an average maternity-leave of six weeks.

**Exercise 10.8**

A recent drug survey showed an increase in use of drugs and alcohol among local high school students as compared to the national percent. Suppose that a survey of 100 local youths and 100 national youths is conducted to see if the percentage of drug and alcohol use is higher locally than nationally.

**Exercise 10.9**

(*Solution on p. 166.*)

A new SAT study course is tested on 12 individuals. Pre-course and post-course scores are recorded. Of interest is the average increase in SAT scores.

**Exercise 10.10**

University of Michigan researchers reported in the *Journal of the National Cancer Institute* that quitting smoking is especially beneficial for those under age 49. In this American Cancer Society study, the risk (probability) of dying of lung cancer was about the same as for those who had never smoked.

**10.2.1 For each problem below, fill in a hypothesis test solution sheet. The solution sheet is in the Appendix and can be copied. For the online version of the book, it is suggested that you copy the .doc or .pdf files.**

NOTE: If you are using a student-t distribution for a homework problem below, including for paired data, you may assume that the underlying population is normally distributed. (In general, you must first prove that assumption, though.)

**Exercise 10.11**

(*Solution on p. 166.*)

A powder diet is tested on 49 people and a liquid diet is tested on 36 different people. Of interest is whether the liquid diet yields a higher average weight loss than the powder diet. The powder diet group had an average weight loss of 42 pounds with a standard deviation of 12 pounds. The liquid diet group had an average weight loss of 45 pounds with a standard deviation of 14 pounds.

**Exercise 10.12**

The average number of English courses taken in a two-year time period by male and female college students is believed to be about the same. An experiment is conducted and data are collected from 29 males and 16 females. The males took an average of 3 English courses with a standard deviation of 0.8. The females took an average of 4 English courses with a standard deviation of 1.0. Are the averages statistically the same?

**Exercise 10.13**

(Solution on p. 166.)

A study is done to determine if students in the California state university system take longer to graduate than students enrolled in private universities. 100 students from both the California state university system and private universities are surveyed. Suppose that from years of research, it is known that the population standard deviations are 1.5811 years and 1 year, respectively. The following data are collected. The California state university system students took on average 4.5 years with a standard deviation of 0.8. The private university students took on average 4.1 years with a standard deviation of 0.3.

**Exercise 10.14**

A new SAT study course is tested on 12 individuals. Pre-course and post-course scores are recorded. Of interest is the average increase in SAT scores. The following data is collected:

Pre-course score	Post-course score
1200	1300
960	920
1010	1100
840	880
1100	1070
1250	1320
860	860
1330	1370
790	770
990	1040
1110	1200
740	850

**Table 10.1****Exercise 10.15**

(Solution on p. 166.)

A recent drug survey showed an increase in use of drugs and alcohol among local high school seniors as compared to the national percent. Suppose that a survey of 100 local seniors and 100 national seniors is conducted to see if the percentage of drug and alcohol use is higher locally than nationally. Locally, 65 seniors reported using drugs or alcohol within the past month, while 60 national seniors reported using them.

**Exercise 10.16**

A student at a four-year college claims that average enrollment at four-year colleges is higher than at two-year colleges in the United States. Two surveys are conducted. Of the 35 two-year colleges surveyed, the average enrollment was 5068 with a standard deviation of 4777. Of the 35 four-year colleges surveyed, the average enrollment was 5466 with a standard deviation of 8191. (Source: Microsoft Bookshelf)

**Exercise 10.17**

(Solution on p. 166.)

A study was conducted by the U.S. Army to see if applying antiperspirant to soldiers' feet for a few days before a major hike would help cut down on the number of blisters soldiers had on their feet. In the experiment, for three nights before they went on a 13-mile hike, a group of 328 West Point cadets put an alcohol-based antiperspirant on their feet. A "control group" of 339 soldiers put on a similar, but inactive, preparation on their feet. On the day of the hike, the temperature reached 83 ° F. At the end of the hike, 21% of the soldiers who had used the antiperspirant and 48% of the control group had developed foot blisters. Conduct a hypothesis test to see if the percent of soldiers using the antiperspirant was significantly lower than the control group. (Source: U.S. Army study reported in *Journal of the American Academy of Dermatologists*)

**Exercise 10.18**

We are interested in whether the percents of female suicide victims for ages 15 to 24 are the same for the white and the black races in the United States. We randomly pick one year, 1992, to compare the races. The number of suicides estimated in the United States in 1992 for white females is 4930. 580 were aged 15 to 24. The estimate for black females is 330. 40 were aged 15 to 24. We will let female suicide victims be our population. (Source: *the National Center for Health Statistics, U.S. Dept. of Health and Human Services*)

**Exercise 10.19**

(Solution on p. 166.)

At Rachel's 11th birthday party, 8 girls were timed to see how long (in seconds) they could hold their breath in a relaxed position. After a two-minute rest, they timed themselves while jumping. The girls thought that the jumping would not affect their times, on average. Test their hypothesis.

Relaxed time (seconds)	Jumping time (seconds)
26	21
47	40
30	28
22	21
23	25
45	43
37	35
29	32

**Table 10.2****Exercise 10.20**

Elizabeth Mjelde, an art history professor, was interested in whether the value from the Golden Ratio formula,  $\left(\frac{\text{larger}+\text{smaller dimension}}{\text{larger dimension}}\right)$  was the same in the Whitney Exhibit for works from 1900 – 1919 as for works from 1920 – 1942. 37 early works were sampled. They averaged 1.74 with a standard deviation of 0.11. 65 of the later works were sampled. They averaged 1.746 with a standard deviation of 0.1064. Do you think that there is a significant difference in the Golden Ratio calculation? (Source: *data from Whitney Exhibit on loan to San Jose Museum of Art*)

**Exercise 10.21**

(Solution on p. 166.)

One of the questions in a study of marital satisfaction of dual-career couples was to rate the statement, "I'm pleased with the way we divide the responsibilities for childcare." The ratings went from 1 (strongly agree) to 5 (strongly disagree). Below are ten of the paired responses for husbands and wives. Conduct a hypothesis test to see if the average difference in the husband's versus the

wife's satisfaction level is negative (meaning that, within the partnership, the husband is happier than the wife).

Wife's score	2	2	3	3	4	2	1	1	2	4
Husband's score	2	2	1	3	2	1	1	1	2	4

Table 10.3

#### Exercise 10.22

Ten individuals went on a low-fat diet for 12 weeks to lower their cholesterol. Evaluate the data below. Do you think that their cholesterol levels were significantly lowered?

Starting cholesterol level	Ending cholesterol level
140	140
220	230
110	120
240	220
200	190
180	150
190	200
360	300
280	300
260	240

Table 10.4

#### Exercise 10.23

(Solution on p. 166.)

Average entry level salaries for college graduates with mechanical engineering degrees and electrical engineering degrees are believed to be approximately the same. (Source: <http://www.graduatingengineer.com><sup>3</sup>). A recruiting office thinks that the average mechanical engineering salary is actually lower than the average electrical engineering salary. The recruiting office randomly surveys 50 entry level mechanical engineers and 60 entry level electrical engineers. Their average salaries were \$46,100 and \$46,700, respectively. Their standard deviations were \$3450 and \$4210, respectively. Conduct a hypothesis test to determine if you agree that the average entry level mechanical engineering salary is lower than the average entry level electrical engineering salary.

#### Exercise 10.24

A recent year was randomly picked from 1985 to the present. In that year, there were 2051 Hispanic students at Cabrillo College out of a total of 12,328 students. At Lake Tahoe College, there were 321 Hispanic students out of a total of 2441 students. In general, do you think that the percent of Hispanic students at the two colleges is basically the same or different? (Source: *Chancellor's Office, California Community Colleges, November 1994*)

#### Exercise 10.25

(Solution on p. 167.)

Eight runners were convinced that the average difference in their individual times for running one mile versus race walking one mile was at most 2 minutes. Below are their times. Do you agree that the average difference is at most 2 minutes?

<sup>3</sup><http://www.graduatingengineer.com/>

Running time (minutes)	Race walking time (minutes)
5.1	7.3
5.6	9.2
6.2	10.4
4.8	6.9
7.1	8.9
4.2	9.5
6.1	9.4
4.4	7.9

Table 10.5

**Exercise 10.26**

Marketing companies have collected data implying that teenage girls use more ring tones on their cellular phones than teenage boys do. In one particular study of 40 randomly chosen teenage girls and boys (20 of each) with cellular phones, the average number of ring tones for the girls was 3.2 with a standard deviation of 1.5. The average for the boys was 1.7 with a standard deviation of 0.8. Conduct a hypothesis test to determine if the averages are approximately the same or if the girls' average is higher than the boys' average.

**Exercise 10.27**

(Solution on p. 167.)

While her husband spent 2½ hours picking out new speakers, a statistician decided to determine whether the percent of men who enjoy shopping for electronic equipment is higher than the percent of women who enjoy shopping for electronic equipment. The population was Saturday afternoon shoppers. Out of 67 men, 24 said they enjoyed the activity. 8 of the 24 women surveyed claimed to enjoy the activity. Interpret the results of the survey.

**Exercise 10.28**

We are interested in whether children's educational computer software costs less, on average, than children's entertainment software. 36 educational software titles were randomly picked from a catalog. The average cost was \$31.14 with a standard deviation of \$4.69. 35 entertainment software titles were randomly picked from the same catalog. The average cost was \$33.86 with a standard deviation of \$10.87. Decide whether children's educational software costs less, on average, than children's entertainment software. (Source: *Educational Resources*, December catalog)

**Exercise 10.29**

(Solution on p. 167.)

Parents of teenage boys often complain that auto insurance costs more, on average, for teenage boys than for teenage girls. A group of concerned parents examines a random sample of insurance bills. The average annual cost for 36 teenage boys was \$679. For 23 teenage girls, it was \$559. From past years, it is known that the population standard deviation for each group is \$180. Determine whether or not you believe that the average cost for auto insurance for teenage boys is greater than that for teenage girls.

**Exercise 10.30**

A group of transfer bound students wondered if they will spend the same average amount on texts and supplies each year at their four-year university as they have at their community college. They conducted a random survey of 54 students at their community college and 66 students at their local four-year university. The sample means were \$947 and \$1011, respectively. The population standard deviations are known to be \$254 and \$87, respectively. Conduct a hypothesis test to determine if the averages are statistically the same.

**Exercise 10.31***(Solution on p. 167.)*

Joan Nguyen recently claimed that the proportion of college-age males with at least one pierced ear is as high as the proportion of college-age females. She conducted a survey in her classes. Out of 107 males, 20 had at least one pierced ear. Out of 92 females, 47 had at least one pierced ear. Do you believe that the proportion of males has reached the proportion of females?

**Exercise 10.32**

Some manufacturers claim that non-hybrid sedan cars have a lower average miles per gallon (mpg) than hybrid ones. Suppose that consumers test 21 hybrid sedans and get an average 31 mpg with a standard deviation of 7 mpg. Thirty-one non-hybrid sedans average 22 mpg with a standard deviation of 4 mpg. Suppose that the population standard deviations are known to be 6 and 3, respectively. Conduct a hypothesis test to the manufacturers claim.

Questions Exercise 10.33 – Exercise 10.37 refer to the Terri Vogel's data set (see Table of Contents).

**Exercise 10.33***(Solution on p. 167.)*

Using the data from Lap 1 only, conduct a hypothesis test to determine if the average time for completing a lap in races is the same as it is in practices.

**Exercise 10.34**

Repeat the test in Exercise 10.33, but use Lap 5 data this time.

**Exercise 10.35***(Solution on p. 167.)*

Repeat the test in Exercise 10.33, but this time combine the data from Laps 1 and 5.

**Exercise 10.36**

In 2 – 3 complete sentences, explain in detail how you might use Terri Vogel's data to answer the following question. "Does Terri Vogel drive faster in races than she does in practices?"

**Exercise 10.37***(Solution on p. 167.)*

Is the proportion of race laps Terri completes slower than 130 seconds less than the proportion of practice laps she completes slower than 135 seconds?

**Exercise 10.38**

"To Breakfast or Not to Breakfast?" by Richard Ayore

In the American society, birthdays are one of those days that everyone looks forward to. People of different ages and peer groups gather to mark the 18th, 20th, . . . birthdays. During this time, one looks back to see what he or she had achieved for the past year, and also focuses ahead for more to come.

If, by any chance, I am invited to one of these parties, my experience is always different. Instead of dancing around with my friends while the music is booming, I get carried away by memories of my family back home in Kenya. I remember the good times I had with my brothers and sister while we did our daily routine.

Every morning, I remember we went to the shamba (garden) to weed our crops. I remember one day arguing with my brother as to why he always remained behind just to join us an hour later. In his defense, he said that he preferred waiting for breakfast before he came to weed. He said, "This is why I always work more hours than you guys!"

And so, to prove his wrong or right, we decided to give it a try. One day we went to work as usual without breakfast, and recorded the time we could work before getting tired and stopping. On the next day, we all ate breakfast before going to work. We recorded how long we worked again before getting tired and stopping. Of interest was our average increase in work time. Though not sure, my brother insisted that it is more than two hours. Using the data below, solve our problem.

Work hours with breakfast	Work hours without breakfast
8	6
7	5
9	5
5	4
9	7
8	7
10	7
7	5
6	6
9	5

Table 10.6

### 10.2.2 Try these multiple choice questions.

For questions Exercise 10.39 – Exercise 10.40, use the following information.

A new AIDS prevention drugs was tried on a group of 224 HIV positive patients. Forty-five (45) patients developed AIDS after four years. In a control group of 224 HIV positive patients, 68 developed AIDS after four years. We want to test whether the method of treatment reduces the proportion of patients that develop AIDS after four years or if the proportions of the treated group and the untreated group stay the same.

Let the subscript  $t$ = treated patient and  $ut$ = untreated patient.

#### Exercise 10.39

(*Solution on p. 167.*)

The appropriate hypotheses are:

- A.  $H_0 : p_t < p_{ut}$  and  $H_a : p_t \geq p_{ut}$
- B.  $H_0 : p_t \leq p_{ut}$  and  $H_a : p_t > p_{ut}$
- C.  $H_0 : p_t = p_{ut}$  and  $H_a : p_t \neq p_{ut}$
- D.  $H_0 : p_t = p_{ut}$  and  $H_a : p_t < p_{ut}$

#### Exercise 10.40

(*Solution on p. 167.*)

If the  $p$ -value is 0.0062 what is the conclusion (use  $\alpha = 5$ )?

- A. The method has no effect.
- B. The method reduces the proportion of HIV positive patients that develop AIDS after four years.
- C. The method increases the proportion of HIV positive patients that develop AIDS after four years.
- D. The test does not determine whether the method helps or does not help.

#### Exercise 10.41

(*Solution on p. 167.*)

Lesley E. Tan investigated the relationship between left-handedness and right-handedness and motor competence in preschool children. Random samples of 41 left-handers and 41 right-handers were given several tests of motor skills to determine if there is evidence of a difference between the children based on this experiment. The experiment produced the means and standard deviations shown below. Determine the appropriate test and best distribution to use for that test.

	Left-handed	Right-handed
Sample size	41	41
Sample mean	97.5	98.1
Sample standard deviation	17.5	19.2

**Table 10.7**

- A. Two independent means, normal distribution
- B. Two independent means, student-t distribution
- C. Matched or paired samples, student-t distribution
- D. Two population proportions, normal distribution

For questions Exercise 10.42 – Exercise 10.43, use the following information.

An experiment is conducted to show that blood pressure can be consciously reduced in people trained in a “biofeedback exercise program.” Six (6) subjects were randomly selected and the blood pressure measurements were recorded before and after the training. The difference between blood pressures was calculated (after – before) producing the following results:  $\bar{x}_d = -10.2$   $s_d = 8.4$ . Using the data, test the hypothesis that the blood pressure has decreased after the training,

**Exercise 10.42***(Solution on p. 167.)*

The distribution for the test is

- A.  $t_5$
- B.  $t_6$
- C.  $N(-10.2, 8.4)$
- D.  $N\left(-10.2, \frac{8.4}{\sqrt{6}}\right)$

**Exercise 10.43***(Solution on p. 167.)*

If  $\alpha = 0.05$ , the  $p$ -value and the conclusion are

- A. 0.0014; the blood pressure decreased after the training
- B. 0.0014; the blood pressure increased after the training
- C. 0.0155; the blood pressure decreased after the training
- D. 0.0155; the blood pressure increased after the training

For questions Exercise 10.44– Exercise 10.45, use the following information.

The Eastern and Western Major League Soccer conferences have a new Reserve Division that allows new players to develop their skills. As of May 25, 2005, the Reserve Division teams scored the following number of goals for 2005.

Western	Eastern
Los Angeles 9	D.C. United 9
FC Dallas 3	Chicago 8
Chivas USA 4	Columbus 7
Real Salt Lake 3	New England 6
Colorado 4	MetroStars 5
San Jose 4	Kansas City 3

**Table 10.8**

Conduct a hypothesis test to determine if the Western Reserve Division teams score, on average, fewer goals than the Eastern Reserve Division teams. Subscripts: 1 Western Reserve Division (W); 2 Eastern Reserve Division (E)

**Exercise 10.44**

(Solution on p. 167.)

The **exact** distribution for the hypothesis test is:

- A. The normal distribution.
- B. The student-t distribution.
- C. The uniform distribution.
- D. The exponential distribution.

**Exercise 10.45**

(Solution on p. 168.)

If the level of significance is 0.05, the conclusion is:

- A. The W Division teams score, on average, fewer goals than the E teams.
- B. The W Division teams score, on average, more goals than the E teams.
- C. The W teams score, on average, about the same number of goals as the E teams score.
- D. Unable to determine.

Questions Exercise 10.46 – Exercise 10.48 refer to the following.

A researcher is interested in determining if a certain drug vaccine prevents West Nile disease. The vaccine with the drug is administered to 36 people and another 36 people are given a vaccine that does not contain the drug. Of the group that gets the vaccine with the drug, one (1) gets West Nile disease. Of the group that gets the vaccine without the drug, three (3) get West Nile disease. Conduct a hypothesis test to determine if the proportion of people that get the vaccine without the drug and get West Nile disease is more than the proportion of people that get the vaccine with the drug and get West Nile disease.

- “Drug” subscript: group who get the vaccine with the drug.
- “No Drug” subscript: group who get the vaccine without the drug

**Exercise 10.46**

(Solution on p. 168.)

This is a test of:

- A. a test of two proportions
- B. a test of two independent means
- C. a test of a single mean
- D. a test of matched pairs.

**Exercise 10.47**

(Solution on p. 168.)

An appropriate null hypothesis is:

- A.  $p_{\text{No Drug}} \leq p_{\text{Drug}}$
- B.  $p_{\text{No Drug}} \geq p_{\text{Drug}}$
- C.  $\mu_{\text{No Drug}} \leq \mu_{\text{Drug}}$
- D.  $p_{\text{No Drug}} > p_{\text{Drug}}$

**Exercise 10.48**

(Solution on p. 168.)

The  $p$ -value is 0.1517. At a 1% level of significance, the appropriate conclusion is

- A. the proportion of people that get the vaccine without the drug and get West Nile disease is less than the proportion of people that get the vaccine with the drug and get West Nile disease.

- B. the proportion of people that get the vaccine without the drug and get West Nile disease is more than the proportion of people that get the vaccine with the drug and get West Nile disease.
- C. the proportion of people that get the vaccine without the drug and get West Nile disease is more than or equal to the proportion of people that get the vaccine with the drug and get West Nile disease.
- D. the proportion of people that get the vaccine without the drug and get West Nile disease is no more than the proportion of people that get the vaccine with the drug and get West Nile disease.

Questions Exercise 10.49 and Exercise 10.50 refer to the following:

A golf instructor is interested in determining if her new technique for improving players' golf scores is effective. She takes four (4) new students. She records their 18-holes scores before learning the technique and then after having taken her class. She conducts a hypothesis test. The data are as follows.

	Player 1	Player 2	Player 3	Player 4
Average score before class	83	78	93	87
Average score after class	80	80	86	86

**Table 10.9**

**Exercise 10.49**

(*Solution on p. 168.*)

This is a test of:

- A. a test of two independent means
- B. a test of two proportions
- C. a test of a single proportion
- D. a test of matched pairs.

**Exercise 10.50**

(*Solution on p. 168.*)

The correct decision is:

- A. Reject  $H_0$
- B. Do not reject  $H_0$
- C. The test is inconclusive

Questions Exercise 10.51 and Exercise 10.52 refer to the following:

Suppose a statistics instructor believes that there is no significant difference between the average class scores of statistics day students on Exam 2 and statistics night students on Exam 2. She takes random samples from each of the populations. The average and standard deviation for 35 statistics day students were 75.86 and 16.91. The average and standard deviation for 37 statistics night students were 75.41 and 19.73. The "day" subscript refers to the statistics day students. The "night" subscript refers to the statistics night students.

**Exercise 10.51**

(*Solution on p. 168.*)

An appropriate alternate hypothesis for the hypothesis test is:

- A.  $\mu_{\text{day}} > \mu_{\text{night}}$
- B.  $\mu_{\text{day}} < \mu_{\text{night}}$
- C.  $\mu_{\text{day}} = \mu_{\text{night}}$

- D.  $\mu_{\text{day}} \neq \mu_{\text{night}}$

**Exercise 10.52**

(*Solution on p. 168.*)

A concluding statement is:

- A. The statistics night students average on Exam 2 is better than the statistics day students average on Exam 2.
- B. The statistics day students average on Exam 2 is better than the statistics night students average on Exam 2.
- C. There is no significant difference between the averages of the statistics day students and night students on Exam 2.
- D. There is a significant difference between the averages of the statistics day students and night students on Exam 2.

### 10.3 Review Questions<sup>4</sup>

The next three questions refer to the following information:

In a survey at Kirkwood Ski Resort the following information was recorded:

**Sport Participation by Age**

	0 – 10	11 – 20	21 – 40	40+
Ski	10	12	30	8
Snowboard	6	17	12	5

**Table 10.10**

Suppose that one person from of the above was randomly selected.

**Exercise 10.53: REVIEW QUESTION 1**

(*Solution on p. 168.*)

Find the probability that the person was a skier or was age 11 – 20.

**Exercise 10.54: REVIEW QUESTION 2**

(*Solution on p. 168.*)

Find the probability that the person was a snowboarder given he/she was age 21 – 40.

**Exercise 10.55: REVIEW QUESTION 3**

(*Solution on p. 168.*)

Explain which of the following are true and which are false.

- a. Sport and Age are independent events.
- b. Ski and age 11 – 20 are mutually exclusive events.
- c.  $P(\text{Ski and age } 21 - 40) < P(\text{Ski} | \text{age } 21 - 40)$
- d.  $P(\text{Snowboard or age } 0 - 10) < P(\text{Snowboard} | \text{age } 0 - 10)$

**Exercise 10.56: REVIEW QUESTION 4**

(*Solution on p. 168.*)

The average length of time a person with a broken leg wears a cast is approximately 6 weeks. The standard deviation is about 3 weeks. Thirty people who had recently healed from broken legs were interviewed. State the distribution that most accurately reflects total time to heal for the thirty people.

---

<sup>4</sup>This content is available online at <<http://cnx.org/content/m19028/1.1/>>.

**Exercise 10.57: REVIEW QUESTION 5**

(Solution on p. 168.)

The distribution for  $X$  is Uniform. What can we say for certain about the distribution for  $\bar{X}$  when  $n = 1$ ?

- A. The distribution for  $\bar{X}$  is still Uniform with the same mean and standard dev. as the distribution for  $X$ .
- B. The distribution for  $\bar{X}$  is Normal with the different mean and a different standard deviation as the distribution for  $X$ .
- C. The distribution for  $\bar{X}$  is Normal with the same mean but a larger standard deviation than the distribution for  $X$ .
- D. The distribution for  $\bar{X}$  is Normal with the same mean but a smaller standard deviation than the distribution for  $X$ .

**Exercise 10.58: REVIEW QUESTION 6**

(Solution on p. 168.)

The distribution for  $X$  is uniform. What can we say for certain about the distribution for  $\sum X$  when  $n = 50$ ?

- A. The distribution for  $\sum X$  is still uniform with the same mean and standard deviation as the distribution for  $X$ .
- B. The distribution for  $\sum X$  is Normal with the same mean but a larger standard deviation as the distribution for  $X$ .
- C. The distribution for  $\sum X$  is Normal with a larger mean and a larger standard deviation than the distribution for  $X$ .
- D. The distribution for  $\sum X$  is Normal with the same mean but a smaller standard deviation than the distribution for  $X$ .

The next three questions refer to the following information:

A group of students measured the lengths of all the carrots in a five-pound bag of baby carrots. They calculated the average length of baby carrots to be 2.0 inches with a standard deviation of 0.25 inches. Suppose we randomly survey 16 five-pound bags of baby carrots.

**Exercise 10.59: REVIEW QUESTION 7**

(Solution on p. 168.)

State the approximate distribution for  $\bar{X}$ , the distribution for the average lengths of baby carrots in 16 five-pound bags.  $\bar{X} \sim$

**Exercise 10.60: REVIEW QUESTION 8**

(Solution on p. 168.)

Explain why we cannot find the probability that one individual randomly chosen carrot is greater than 2.25 inches.

**Exercise 10.61: REVIEW QUESTION 9**

(Solution on p. 168.)

Find the probability that  $\bar{X}$  is between 2 and 2.25 inches.

The next three questions refer to the following information:

At the beginning of the term, the amount of time a student waits in line at the campus store is normally distributed with a mean of 5 minutes and a standard deviation of 2 minutes.

**Exercise 10.62: REVIEW QUESTION 10**

(Solution on p. 168.)

Find the 90th percentile of waiting time in minutes.

**Exercise 10.63: REVIEW QUESTION 11**

(Solution on p. 168.)

Find the median waiting time for one student.

**Exercise 10.64: REVIEW QUESTION 12**

(Solution on p. 168.)

Find the probability that the average waiting time for 40 students is at least 4.5 minutes.

## Solutions to Exercises in Chapter 10

### Solution to Exercise 10.1 (p. 153)

A

### Solution to Exercise 10.3 (p. 154)

B

### Solution to Exercise 10.5 (p. 154)

A

### Solution to Exercise 10.7 (p. 154)

D

### Solution to Exercise 10.9 (p. 154)

C

### Solution to Exercise 10.11 (p. 154)

d.  $t_{68.44}$ 

e. -1.04

f. 0.1519

h. Dec: do not reject null

### Solution to Exercise 10.13 (p. 155)

Standard Normal

e.  $z = 2.14$ 

f. 0.0163

h. Decision: Reject null when  $\alpha = 0.05$ ; Do not reject null when  $\alpha = 0.01$ 

### Solution to Exercise 10.15 (p. 155)

e. 0.73

f. 0.2326

h. Decision: Do not reject null

### Solution to Exercise 10.17 (p. 156)

e. -7.33

f. 0

h. Decision: Reject null

### Solution to Exercise 10.19 (p. 156)

d.  $t_7$ 

e. -1.51

f. 0.1755

h. Decision: Do not reject null

### Solution to Exercise 10.21 (p. 156)

d.  $t_9$ e.  $t = -1.86$ 

f. 0.0479

h. Decision: Reject null, but run another test

### Solution to Exercise 10.23 (p. 157)

d.  $t_{108}$ e.  $t = -0.82$ 

f. 0.2066

h. Decision: Do not reject null

**Solution to Exercise 10.25 (p. 157)**

d.  $t_7$

e.  $t = 2.9850$

f. 0.0103

h. Decision: Reject null; The average difference is more than 2 minutes.

**Solution to Exercise 10.27 (p. 158)**

e. 0.22

f. 0.4133

h. Decision: Do not reject null

**Solution to Exercise 10.29 (p. 158)**

e.  $z = 2.50$

f. 0.0063

h. Decision: Reject null

**Solution to Exercise 10.31 (p. 159)**

e. -4.82

f. 0

h. Decision: Reject null

**Solution to Exercise 10.33 (p. 159)**

d.  $t_{20.32}$

e. -4.70

f. 0.0001

h. Decision: Reject null

**Solution to Exercise 10.35 (p. 159)**

d.  $t_{40.94}$

e. -5.08

f. 0

h. Decision: Reject null

**Solution to Exercise 10.37 (p. 159)**

e. -0.95

f. 0.1705

h. Decision: Do not reject null

**Solution to Exercise 10.39 (p. 160)**

D

**Solution to Exercise 10.40 (p. 160)**

B

**Solution to Exercise 10.41 (p. 160)**

B

**Solution to Exercise 10.42 (p. 161)**

A

**Solution to Exercise 10.43 (p. 161)**

C

**Solution to Exercise 10.44 (p. 162)**

B

**Solution to Exercise 10.45 (p. 162)**

C

**Solution to Exercise 10.46 (p. 162)**

A

**Solution to Exercise 10.47 (p. 162)**

A

**Solution to Exercise 10.48 (p. 162)**

D

**Solution to Exercise 10.49 (p. 163)**

D

**Solution to Exercise 10.50 (p. 163)**

B

**Solution to Exercise 10.51 (p. 163)**

D

**Solution to Exercise 10.52 (p. 164)**

C

**Solution to Exercise 10.53 (p. 164)**REVIEW QUESTION 1 Solution :  $\frac{77}{100}$ **Solution to Exercise 10.54 (p. 164)**REVIEW QUESTION 2 Solution :  $\frac{12}{42}$ **Solution to Exercise 10.55 (p. 164)****REVIEW QUESTION 3 Solution**

- a. False
- b. False
- c. True
- d. False

**Solution to Exercise 10.56 (p. 164)**REVIEW QUESTION 4 Solution :  $N(180, 16.43)$ **Solution to Exercise 10.57 (p. 165)**

REVIEW QUESTION 5 Solution : A

**Solution to Exercise 10.58 (p. 165)**

REVIEW QUESTION 6 Solution : C

**Solution to Exercise 10.59 (p. 165)**REVIEW QUESTION 7 Solution :  $N\left(2, \frac{25}{\sqrt{16}}\right)$ **Solution to Exercise 10.60 (p. 165)****REVIEW QUESTION 8 Solution**

We do not know the probability distribution for the underlying population of lengths of the individual carrots.

**Solution to Exercise 10.61 (p. 165)**

REVIEW QUESTION 9 Solution : 0.5000

**Solution to Exercise 10.62 (p. 165)**

REVIEW QUESTION 10 Solution : 7.6

**Solution to Exercise 10.63 (p. 165)**

REVIEW QUESTION 11 Solution : 5

**Solution to Exercise 10.64 (p. 165)**

REVIEW QUESTION 12 Solution : 0.9431

# Chapter 11

## The Chi-Square Distribution

### 11.1 Summary of Formulas<sup>1</sup>

**Rule 11.1:** The Chi-square Probability Distribution

$$\mu = df \text{ and } \sigma = \sqrt{2 \cdot df}$$

**Rule 11.2:** Goodness-of-Fit Hypothesis Test

- Use goodness-of-fit to test whether a data set fits a particular probability distribution.
- The degrees of freedom are number of cells or categories - 1.
- The test statistic is  $\sum_n \frac{(O-E)^2}{E}$ , where  $O$  = observed values (data),  $E$  = expected values (from theory), and  $n$  = the number of different data cells or categories.
- The test is right-tailed.

**Rule 11.3:** Test of Independence

- Use the test of independence to test whether two factors are independent or not.
- The degrees of freedom are equal to (number of columns - 1)(number of rows - 1).
- The test statistic is  $\sum_{(i,j)} \frac{(O-E)^2}{E}$  where  $O$  = observed values,  $E$  = expected values,  $i$  = the number of rows in the table, and  $j$  = the number of columns in the table.
- The test is right-tailed.
- If the null hypothesis is true, the expected number  $E = \frac{(\text{row total})(\text{column total})}{\text{total surveyed}}$ .

**Rule 11.4:** Test of a Single Variance

- Use the test to determine variation.
- The degrees of freedom are the number of samples - 1.
- The test statistic is  $\frac{(n-1) \cdot s^2}{\sigma^2}$ , where  $n$  = the total number of data,  $s^2$  = sample variance, and  $\sigma^2$  = population variance.
- The test may be left, right, or two-tailed.

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m17058/1.5/>>.

## 11.2 Homework<sup>2</sup>

### Exercise 11.1

- a. Explain why the “goodness of fit” test and the “test for independence” are generally right tailed tests.
- b. If you did a left-tailed test, what would you be testing?

### 11.2.1 Word Problems

For each word problem, use a solution sheet to solve the hypothesis test problem. Go to The Table of Contents 14. Appendix for the solution sheet. Round expected frequency to two decimal places.

### Exercise 11.2

A 6-sided die is rolled 120 times. Fill in the expected frequency column. Then, conduct a hypothesis test to determine if the die is fair. The data below are the result of the 120 rolls.

Face Value	Frequency	Expected Frequency
1	15	
2	29	
3	16	
4	15	
5	30	
6	15	

Table 11.1

### Exercise 11.3

(*Solution on p. 182.*)

The marital status distribution of the U.S. male population, age 15 and older, is as shown below. (*Source: U.S. Census Bureau, Current Population Reports*)

Marital Status	Percent	Expected Frequency
never married	31.3	
married	56.1	
widowed	2.5	
divorced/separated	10.1	

Table 11.2

Suppose that a random sample of 400 U.S. young adult males, 18 – 24 years old, yielded the following frequency distribution. We are interested in whether this age group of males fits the distribution of the U.S. adult population. Calculate the frequency one would expect when surveying 400 people. Fill in the above table, rounding to two decimal places.

<sup>2</sup>This content is available online at <<http://cnx.org/content/m17028/1.18/>>.

Marital Status	Frequency
never married	140
married	238
widowed	2
divorced/separated	20

**Table 11.3**

**The next two questions refer to the following information.** The columns in the chart below contain the Race/Ethnicity of U.S. Public Schools: High School Class of 2009, the percentages for the Advanced Placement Examinee Population for that class and the Overall Student Population. (Source: <http://www.collegeboard.com>). Suppose the right column contains the result of a survey of 1000 local students from the Class of 2009 who took an AP Exam.

Race/Ethnicity	AP Examinee Population	Overall Student Population	Survey Frequency
Asian, Asian American or Pacific Islander	10.2%	5.4%	113
Black or African American	8.2%	14.5%	94
Hispanic or Latino	15.5%	15.9%	136
American Indian or Alaska Native	0.6%	1.2%	10
White	59.4%	61.6%	604
Not reported/other	6.1%	1.4%	43

**Table 11.4****Exercise 11.4**

Perform a goodness-of-fit test to determine whether the local results follow the distribution of the U. S. Overall Student Population based on ethnicity.

**Exercise 11.5**

(*Solution on p. 182.*)

Perform a goodness-of-fit test to determine whether the local results follow the distribution of U. S. AP Examinee Population, based on ethnicity.

**Exercise 11.6**

The City of South Lake Tahoe, CA, has an Asian population of 1419 people, out of a total population of 23,609 (Source: *U.S. Census Bureau, Census 2000*). Suppose that a survey of 1419 self-reported Asians in Manhattan, NY, area yielded the data in the table below. Conduct a goodness of fit test to determine if the self-reported sub-groups of Asians in the Manhattan area fit that of the Lake Tahoe area.

Race	Lake Tahoe Frequency	Manhattan Frequency	
Asian Indian	131	174	
Chinese	118	557	
Filipino	1045	518	
Japanese	80	54	
Korean	12	29	
Vietnamese	9	21	
Other	24	66	

Table 11.5

**The next two questions refer to the following information:** UCLA conducted a survey of more than 263,000 college freshmen from 385 colleges in fall 2005. The results of student expected majors by gender were reported in *The Chronicle of Higher Education* (2/2/2006). Suppose a survey of 5000 graduating females and 5000 graduating males was done as a follow-up in 2010 to determine what their actual major was. The results are shown in the tables for Exercises 7 and 8. The second column in each table does not add to 100% because of rounding.

**Exercise 11.7**

(Solution on p. 182.)

Conduct a hypothesis test to determine if the actual college major of graduating females fits the distribution of their expected majors.

Major	Women - Expected Major	Women - Actual Major
Arts & Humanities	14.0%	670
Biological Sciences	8.4%	410
Business	13.1%	685
Education	13.0%	650
Engineering	2.6%	145
Physical Sciences	2.6%	125
Professional	18.9%	975
Social Sciences	13.0%	605
Technical	0.4%	15
Other	5.8%	300
Undecided	8.0%	420

Table 11.6

**Exercise 11.8**

Conduct a hypothesis test to determine if the actual college major of graduating males fits the distribution of their expected majors.

Major	Men - Expected Major	Men - Actual Major
Arts & Humanities	11.0%	600
Biological Sciences	6.7%	330
Business	22.7%	1130
Education	5.8%	305
Engineering	15.6%	800
Physical Sciences	3.6%	175
Professional	9.3%	460
Social Sciences	7.6%	370
Technical	1.8%	90
Other	8.2%	400
Undecided	6.6%	340

**Table 11.7****Exercise 11.9***(Solution on p. 182.)*

A recent debate about where in the United States skiers believe the skiing is best prompted the following survey. Test to see if the best ski area is independent of the level of the skier.

U.S. Ski Area	Beginner	Intermediate	Advanced
Tahoe	20	30	40
Utah	10	30	60
Colorado	10	40	50

**Table 11.8****Exercise 11.10**

Car manufacturers are interested in whether there is a relationship between the size of car an individual drives and the number of people in the driver's family (that is, whether car size and family size are independent). To test this, suppose that 800 car owners were randomly surveyed with the following results. Conduct a test for independence.

Family Size	Sub & Compact	Mid-size	Full-size	Van & Truck
1	20	35	40	35
2	20	50	70	80
3 - 4	20	50	100	90
5+	20	30	70	70

**Table 11.9****Exercise 11.11***(Solution on p. 182.)*

College students may be interested in whether or not their majors have any effect on starting salaries after graduation. Suppose that 300 recent graduates were surveyed as to their majors

in college and their starting salaries after graduation. Below are the data. Conduct a test for independence.

Major	< \$30,000	\$30,000 - \$39,999	\$40,000 +
English	5	20	5
Engineering	10	30	60
Nursing	10	15	15
Business	10	20	30
Psychology	20	30	20

Table 11.10

**Exercise 11.12**

Some travel agents claim that honeymoon hot spots vary according to age of the bride and groom. Suppose that 280 East Coast recent brides were interviewed as to where they spent their honeymoons. The information is given below. Conduct a test for independence.

Location	20 - 29	30 - 39	40 - 49	50 and over
Niagara Falls	15	25	25	20
Poconos	15	25	25	10
Europe	10	25	15	5
Virgin Islands	20	25	15	5

Table 11.11

**Exercise 11.13**

(Solution on p. 182.)

A manager of a sports club keeps information concerning the main sport in which members participate and their ages. To test whether there is a relationship between the age of a member and his or her choice of sport, 643 members of the sports club are randomly selected. Conduct a test for independence.

Sport	18 - 25	26 - 30	31 - 40	41 and over
racquetball	42	58	30	46
tennis	58	76	38	65
swimming	72	60	65	33

Table 11.12

**Exercise 11.14**

A major food manufacturer is concerned that the sales for its skinny French fries have been decreasing. As a part of a feasibility study, the company conducts research into the types of fries sold across the country to determine if the type of fries sold is independent of the area of the country. The results of the study are below. Conduct a test for independence.

Type of Fries	Northeast	South	Central	West
skinny fries	70	50	20	25
curly fries	100	60	15	30
steak fries	20	40	10	10

**Table 11.13****Exercise 11.15***(Solution on p. 182.)*

According to Dan Lenard, an independent insurance agent in the Buffalo, N.Y. area, the following is a breakdown of the amount of life insurance purchased by males in the following age groups. He is interested in whether the age of the male and the amount of life insurance purchased are independent events. Conduct a test for independence.

Age of Males	None	\$50,000 - \$100,000	\$100,001 - \$150,000	\$150,001 - \$200,000	\$200,000 +
20 - 29	40	15	40	0	5
30 - 39	35	5	20	20	10
40 - 49	20	0	30	0	30
50 +	40	30	15	15	10

**Table 11.14****Exercise 11.16**

Suppose that 600 thirty-year-olds were surveyed to determine whether or not there is a relationship between the level of education an individual has and salary. Conduct a test for independence.

Annual Salary	Not a high school grad.	High school graduate	College graduate	Masters or doctorate
< \$30,000	15	25	10	5
\$30,000 - \$40,000	20	40	70	30
\$40,000 - \$50,000	10	20	40	55
\$50,000 - \$60,000	5	10	20	60
\$60,000 +	0	5	10	150

**Table 11.15****Exercise 11.17***(Solution on p. 182.)*

A plant manager is concerned her equipment may need recalibrating. It seems that the actual weight of the 15 oz. cereal boxes it fills has been fluctuating. The standard deviation should be at most  $\frac{1}{2}$  oz. In order to determine if the machine needs to be recalibrated, 84 randomly selected boxes of cereal from the next day's production were weighed. The standard deviation of the 84 boxes was 0.54. Does the machine need to be recalibrated?

**Exercise 11.18**

Consumers may be interested in whether the cost of a particular calculator varies from store to store. Based on surveying 43 stores, which yielded a sample mean of \$84 and a sample standard deviation of \$12, test the claim that the standard deviation is greater than \$15.

**Exercise 11.19**

(Solution on p. 183.)

Isabella, an accomplished **Bay to Breakers** runner, claims that the standard deviation for her time to run the  $7 \frac{1}{2}$  mile race is at most 3 minutes. To test her claim, Rupinder looks up 5 of her race times. They are 55 minutes, 61 minutes, 58 minutes, 63 minutes, and 57 minutes.

**Exercise 11.20**

Airline companies are interested in the consistency of the number of babies on each flight, so that they have adequate safety equipment. They are also interested in the variation of the number of babies. Suppose that an airline executive believes the average number of babies on flights is 6 with a variance of 9 at most. The airline conducts a survey. The results of the 18 flights surveyed give a sample average of 6.4 with a sample standard deviation of 3.9. Conduct a hypothesis test of the airline executive's belief.

**Exercise 11.21**

(Solution on p. 183.)

According to the *U.S. Bureau of the Census, United Nations*, in 1994 the number of births per woman in China was 1.8. This fertility rate has been attributed to the law passed in 1979 restricting births to one per woman. Suppose that a group of students studied whether or not the standard deviation of births per woman was greater than 0.75. They asked 50 women across China the number of births they had. Below are the results. Does the students' survey indicate that the standard deviation is greater than 0.75?

# of births	Frequency
0	5
1	30
2	10
3	5

**Table 11.16****Exercise 11.22**

According to an avid aquarist, the average number of fish in a 20-gallon tank is 10, with a standard deviation of 2. His friend, also an aquarist, does not believe that the standard deviation is 2. She counts the number of fish in 15 other 20-gallon tanks. Based on the results that follow, do you think that the standard deviation is different from 2? Data: 11; 10; 9; 10; 10; 11; 11; 10; 12; 9; 7; 9; 11; 10; 11

**Exercise 11.23**

(Solution on p. 183.)

The manager of "Frenchies" is concerned that patrons are not consistently receiving the same amount of French fries with each order. The chef claims that the standard deviation for a 10-ounce order of fries is at most 1.5 oz., but the manager thinks that it may be higher. He randomly weighs 49 orders of fries, which yields: mean of 11 oz., standard deviation of 2 oz.

**11.2.2 Try these true/false questions.****Exercise 11.24**

(Solution on p. 183.)

As the degrees of freedom increase, the graph of the chi-square distribution looks more and more symmetrical.

**Exercise 11.25**

(Solution on p. 183.)

The standard deviation of the chi-square distribution is twice the mean.

**Exercise 11.26** *(Solution on p. 183.)*

The mean and the median of the chi-square distribution are the same if  $df = 24$ .

**Exercise 11.27** *(Solution on p. 183.)*

In a Goodness-of-Fit test, the expected values are the values we would expect if the null hypothesis were true.

**Exercise 11.28** *(Solution on p. 183.)*

In general, if the observed values and expected values of a Goodness-of-Fit test are not close together, then the test statistic can get very large and on a graph will be way out in the right tail.

**Exercise 11.29** *(Solution on p. 183.)*

The degrees of freedom for a Test for Independence are equal to the sample size minus 1.

**Exercise 11.30** *(Solution on p. 183.)*

Use a Goodness-of-Fit test to determine if high school principals believe that students are absent equally during the week or not.

**Exercise 11.31** *(Solution on p. 183.)*

The Test for Independence uses tables of observed and expected data values.

**Exercise 11.32** *(Solution on p. 183.)*

The test to use when determining if the college or university a student chooses to attend is related to his/her socioeconomic status is a Test for Independence.

**Exercise 11.33** *(Solution on p. 183.)*

The test to use to determine if a six-sided die is fair is a Goodness-of-Fit test.

**Exercise 11.34** *(Solution on p. 183.)*

In a Test of Independence, the expected number is equal to the row total multiplied by the column total divided by the total surveyed.

**Exercise 11.35** *(Solution on p. 183.)*

In a Goodness-of Fit test, if the p-value is 0.0113, in general, do not reject the null hypothesis.

**Exercise 11.36** *(Solution on p. 184.)*

For a Chi-Square distribution with degrees of freedom of 17, the probability that a value is greater than 20 is 0.7258.

**Exercise 11.37** *(Solution on p. 184.)*

If  $df = 2$ , the chi-square distribution has a shape that reminds us of the exponential.

## 11.3 Review Questions<sup>3</sup>

**The next two questions refer to the following real study:**

A recent survey of U.S. teenage pregnancy was answered by 720 girls, age 12 - 19. 6% of the girls surveyed said they have been pregnant. (*Parade Magazine*) We are interested in the true proportion of U.S. girls, age 12 - 19, who have been pregnant.

**Exercise 11.38: REVIEW QUESTION 1** *(Solution on p. 184.)*

Find the 95% confidence interval for the true proportion of U.S. girls, age 12 - 19, who have been pregnant.

**Exercise 11.39: REVIEW QUESTION 2** *(Solution on p. 184.)*

The report also stated that the results of the survey are accurate to within  $\pm 3.7\%$  at the 95% confidence level. Suppose that a new study is to be done. It is desired to be accurate to within 2% of the 95% confidence level. What will happen to the minimum number that should be surveyed?

---

<sup>3</sup>This content is available online at <<http://cnx.org/content/m19026/1.1/>>.

**Exercise 11.40: REVIEW QUESTION 3**

Given:  $X \sim \text{Exp}\left(\frac{1}{3}\right)$ . Sketch the graph that depicts:  $P(X > 1)$ .

**The next four questions refer to the following information:**

Suppose that the time that owners keep their cars (purchased new) is normally distributed with a mean of 7 years and a standard deviation of 2 years. We are interested in how long an individual keeps his car (purchased new). Our population is people who buy their cars new.

**Exercise 11.41: REVIEW QUESTION 4**

(*Solution on p. 184.*)

60% of individuals keep their cars **at most** how many years?

**Exercise 11.42: REVIEW QUESTION 5**

(*Solution on p. 184.*)

Suppose that we randomly survey one person. Find the probability that person keeps his/her car **less than** 2.5 years.

**Exercise 11.43: REVIEW QUESTION 6**

(*Solution on p. 184.*)

If we are to pick individuals 10 at a time, find the distribution for the **average** car length ownership.

**Exercise 11.44: REVIEW QUESTION 7**

(*Solution on p. 184.*)

If we are to pick 10 individuals, find the probability that the **sum** of their ownership time is more than 55 years.

**Exercise 11.45: REVIEW QUESTION 8**

(*Solution on p. 184.*)

For which distribution is the median not equal to the mean?

- A. Uniform
- B. Exponential
- C. Normal
- D. Student-t

**Exercise 11.46: REVIEW QUESTION 9**

(*Solution on p. 184.*)

Compare the standard normal distribution to the student-t distribution, centered at 0. Explain which of the following are true and which are false.

- a. As the number surveyed increases, the area to the left of -1 for the student-t distribution approaches the area for the standard normal distribution.
- b. As the number surveyed increases, the area to the left of -1 for the standard normal distribution approaches the area for the student-t distribution.
- c. As the degrees of freedom decrease, the graph of the student-t distribution looks more like the graph of the standard normal distribution.
- d. If the number surveyed is less than 30, the normal distribution should never be used.

**The next five questions refer to the following information:**

We are interested in the checking account balance of a twenty-year-old college student. We randomly survey 16 twenty-year-old college students. We obtain a sample mean of \$640 and a sample standard deviation of \$150. Let  $X$  = checking account balance of an individual twenty year old college student.

**Exercise 11.47: REVIEW QUESTION 10**

Explain why we cannot determine the distribution of  $X$ .

**Exercise 11.48: REVIEW QUESTION 11**

(*Solution on p. 184.*)

If you were to create a confidence interval or perform a hypothesis test for the population average checking account balance of 20-year old college students, what distribution would you use?

**Exercise 11.49: REVIEW QUESTION 12***(Solution on p. 184.)*

Find the 95% confidence interval for the true average checking account balance of a twenty-year-old college student.

**Exercise 11.50: REVIEW QUESTION 13***(Solution on p. 184.)*

What type of data is the balance of the checking account considered to be?

**Exercise 11.51: REVIEW QUESTION 14***(Solution on p. 184.)*

What type of data is the number of 20 year olds considered to be?

**Exercise 11.52: REVIEW QUESTION 15***(Solution on p. 184.)*

On average, a busy emergency room gets a patient with a shotgun wound about once per week. We are interested in the number of patients with a shotgun wound the emergency room gets per 28 days.

- a. Define the random variable  $X$ .
- b. State the distribution for  $X$ .
- c. Find the probability that the emergency room gets no patients with shotgun wounds in the next 28 days.

**The next two questions refer to the following information:**

The probability that a certain slot machine will pay back money when a quarter is inserted is 0.30 . Assume that each play of the slot machine is independent from each other. A person puts in 15 quarters for 15 plays.

**Exercise 11.53: REVIEW QUESTION 16**

REVIEW QUESTION 16 Removed

**Exercise 11.54: REVIEW QUESTION 17***(Solution on p. 184.)*

Is it likely that exactly 8 of the 15 plays would pay back money? Justify your answer numerically.

**Exercise 11.55: REVIEW QUESTION 18***(Solution on p. 184.)*

A game is played with the following rules:

- it costs \$10 to enter
- a fair coin is tossed 4 times
- if you do not get 4 heads or 4 tails, you lose your \$10
- if you get 4 heads or 4 tails, you get back your \$10, plus \$30 more

Over the long run of playing this game, what are your expected earnings?

**Exercise 11.56: REVIEW QUESTION 19***(Solution on p. 184.)*

- The average grade on a math exam in Rachel's class was 74, with a standard deviation of 5. Rachel earned an 80.
- The average grade on a math exam in Becca's class was 47, with a standard deviation of 2. Becca earned a 51.
- The average grade on a math exam in Matt's class was 70, with a standard deviation of 8. Matt earned an 83.

Find whose score was the best, compared to his or her own class. Justify your answer numerically.

**The next two questions refer to the following information:**

70 compulsive gamblers were asked the number of days they go to casinos per week. The results are given in the following graph:

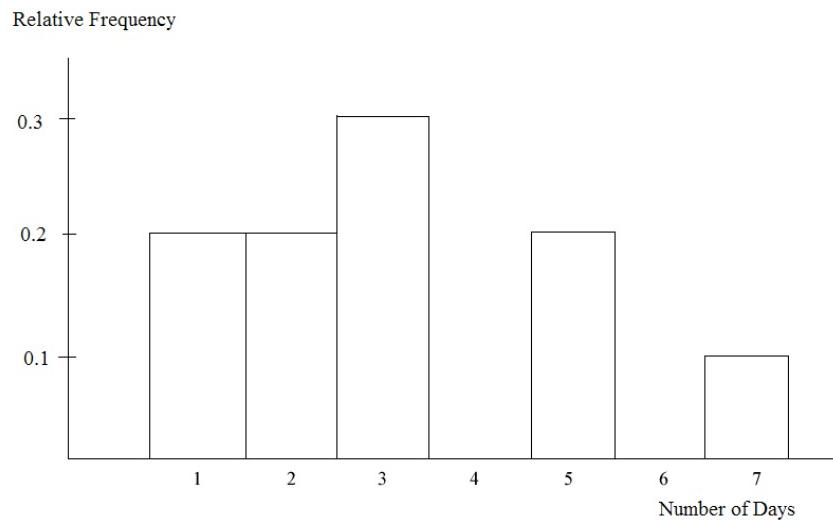


Figure 11.1

**Exercise 11.57: REVIEW QUESTION 20**

(Solution on p. 184.)

Find the number of responses that were "5".

**Exercise 11.58: REVIEW QUESTION 21**

(Solution on p. 184.)

Find the mean, standard deviation, all four quartiles and IQR.

**Exercise 11.59: REVIEW QUESTION 22**

(Solution on p. 185.)

Based upon research at De Anza College, it is believed that about 19% of the student population speaks a language other than English at home.

Suppose that a study was done this year to see if that percent has decreased. Ninety-eight students were randomly surveyed with the following results. Fourteen said that they speak a language other than English at home.

- State an appropriate **null** hypothesis.
- State an appropriate **alternate** hypothesis.
- Define the Random Variable,  $P'$ .
- Calculate the test statistic.
- Calculate the p-value.
- At the 5% level of decision, what is your decision about the null hypothesis?
- What is the Type I error?
- What is the Type II error?

**Exercise 11.60: REVIEW QUESTION 23**

(Solution on p. 185.)

Assume that you are an emergency paramedic called in to rescue victims of an accident. You need to help a patient who is bleeding profusely. The patient is also considered to be a high risk for contracting AIDS. Assume that the null hypothesis is that the patient does not have the HIV virus. What is a Type I error?

**Exercise 11.61: REVIEW QUESTION 24**

(Solution on p. 185.)

It is often said that Californians are more casual than the rest of Americans. Suppose that a survey was done to see if the proportion of Californian professionals that wear jeans to work is

greater than the proportion of non-Californian professionals. Fifty of each was surveyed with the following results. 10 Californians wear jeans to work and 4 non-Californians wear jeans to work.

- $C$  = Californian professional
  - NC = non-Californian professional
- 
- a. State appropriate **null** and **alternate** hypotheses.
  - b. Define the Random Variable.
  - c. Calculate the test statistic and p-value.
  - d. At the 5% level of decision, do you accept or reject the null hypothesis?
  - e. What is the Type I error?
  - f. What is the Type II error?

**The next two questions refer to the following information:**

A group of Statistics students have developed a technique that they feel will lower their anxiety level on statistics exams. They measured their anxiety level at the start of the quarter and again at the end of the quarter. Recorded is the paired data in that order: (1000, 900); (1200, 1050); (600, 700); (1300, 1100); (1000, 900); (900, 900).

**Exercise 11.62: REVIEW QUESTION 25**

*(Solution on p. 185.)*

This is a test of (pick the best answer):

- A. large samples, independent means
- B. small samples, independent means
- C. dependent means

**Exercise 11.63: REVIEW QUESTION 26**

*(Solution on p. 185.)*

State the distribution to use for the test.

## Solutions to Exercises in Chapter 11

### Solution to Exercise 11.3 (p. 170)

- a. The data fits the distribution
- b. The data does not fit the distribution
- c. 3
- e. 19.27
- f. 0.0002
- h. Decision: Reject Null; Conclusion: Data does not fit the distribution.

### Solution to Exercise 11.5 (p. 171)

- c. 5
- e. 13.4
- f. 0.0199
- g. Decision: Reject null when  $\alpha = 0.05$ ; Conclusion: Local data do not fit the AP Examinee Distribution.  
Decision: Do not reject null when  $\alpha = 0.01$ ; Conclusion: Local data do fit the AP Examinee Distribution.

### Solution to Exercise 11.7 (p. 172)

- c. 10
- e. 11.48
- f. 0.3214
- h. Decision: Do not reject null when  $\alpha = 0.05$  and  $\alpha = 0.01$ ; Conclusion: Distribution of majors by graduating females fits the distribution of expected majors.

### Solution to Exercise 11.9 (p. 173)

- c. 4
- e. 10.53
- f. 0.0324
- h. Decision: Reject null; Conclusion: Best ski area and level of skier are not independent.

### Solution to Exercise 11.11 (p. 173)

- c. 8
- e. 33.55
- f. 0
- h. Decision: Reject null; Conclusion: Major and starting salary are not independent events.

### Solution to Exercise 11.13 (p. 174)

- c. 6
- e. 25.21
- f. 0.0003
- h. Decision: Reject null

### Solution to Exercise 11.15 (p. 175)

- c. 12
- e. 125.74
- f. 0
- h. Decision: Reject null

### Solution to Exercise 11.17 (p. 175)

- c. 83
- d. 96.81
- e. 0.1426
- g. Decision: Do not reject null; Conclusion: The standard deviation is at most 0.5 oz.
- h. It does not need to be calibrated

**Solution to Exercise 11.19 (p. 176)**

- c. 4
- d. 4.52
- e. 0.3402
- g. Decision: Do not reject null.
- h. No

**Solution to Exercise 11.21 (p. 176)**

- c. 49
- d. 54.37
- e. 0.2774
- g. Decision: Do not reject null; Conclusion: The standard deviation is at most 0.75.
- h. No

**Solution to Exercise 11.23 (p. 176)**

- a.  $\sigma^2 \leq (1.5)^2$
- c. 48
- d. 85.33
- e. 0.0007
- g. Decision: Reject null.
- h. Yes

**Solution to Exercise 11.24 (p. 176)**

True

**Solution to Exercise 11.25 (p. 176)**

False

**Solution to Exercise 11.26 (p. 177)**

False

**Solution to Exercise 11.27 (p. 177)**

True

**Solution to Exercise 11.28 (p. 177)**

True

**Solution to Exercise 11.29 (p. 177)**

False

**Solution to Exercise 11.30 (p. 177)**

True

**Solution to Exercise 11.31 (p. 177)**

True

**Solution to Exercise 11.32 (p. 177)**

True

**Solution to Exercise 11.33 (p. 177)**

True

**Solution to Exercise 11.34 (p. 177)**

True

**Solution to Exercise 11.35 (p. 177)**

False

**Solution to Exercise 11.36 (p. 177)**

False

**Solution to Exercise 11.37 (p. 177)**

True

**Solution to Exercise 11.38 (p. 177)**

REVIEW QUESTION 1 Solution : (0.0424, 0.0770)

**Solution to Exercise 11.39 (p. 177)**

REVIEW QUESTION 2 Solution : 2401

**Solution to Exercise 11.41 (p. 178)**

REVIEW QUESTION 4 Solution : 7.5 years

**Solution to Exercise 11.42 (p. 178)**

REVIEW QUESTION 5 Solution : 0.0122

**Solution to Exercise 11.43 (p. 178)**REVIEW QUESTION 6 Solution :  $N(7, 0.63)$ **Solution to Exercise 11.44 (p. 178)**

REVIEW QUESTION 7 Solution : 0.9911

**Solution to Exercise 11.45 (p. 178)**

REVIEW QUESTION 8 Solution : B

**Solution to Exercise 11.46 (p. 178)****REVIEW QUESTION 9 Solution**

- a. True
- b. False
- c. False
- d. False

**Solution to Exercise 11.48 (p. 178)**REVIEW QUESTION 11 Solution : student-t with  $df = 15$ **Solution to Exercise 11.49 (p. 179)**

REVIEW QUESTION 12 Solution : (560.07, 719.93)

**Solution to Exercise 11.50 (p. 179)**

REVIEW QUESTION 13 Solution : quantitative - continuous

**Solution to Exercise 11.51 (p. 179)**

REVIEW QUESTION 14 Solution : quantitative - discrete

**Solution to Exercise 11.52 (p. 179)****REVIEW QUESTION 15 Solution**

- b.  $P(4)$
- c. 0.0183

**Solution to Exercise 11.54 (p. 179)**REVIEW QUESTION 17 Solution : No;  $P(X = 8) = 0.0348$ **Solution to Exercise 11.55 (p. 179)**

REVIEW QUESTION 18 Solution : You will lose \$5

**Solution to Exercise 11.56 (p. 179)**

REVIEW QUESTION 19 Solution : Becca

**Solution to Exercise 11.57 (p. 180)**

REVIEW QUESTION 20 Solution : 14

**Solution to Exercise 11.58 (p. 180)****REVIEW QUESTION 21 Solution**

- Mean = 3.2
- Quartiles = 1.85, 2, 3, and 5
- IQR = 3

**Solution to Exercise 11.59 (p. 180)**

**REVIEW QUESTION 22 Solution**

- d.  $z = -1.19$
- e. 0.1171
- f. Do not reject the null

**Solution to Exercise 11.60 (p. 180)**

**REVIEW QUESTION 23 Solution**

The Type I error would be to conclude that the patient has the HIV virus when in reality the patient does not have the HIV virus.

**Solution to Exercise 11.61 (p. 180)**

**REVIEW QUESTION 24 Solution**

- c.  $z = 1.73 ; p = 0.0419$
- d. Reject the null

**Solution to Exercise 11.62 (p. 181)**

**REVIEW QUESTION 25 Solution : C**

**Solution to Exercise 11.63 (p. 181)**

**REVIEW QUESTION 26 Solution :  $t_5$**



## Chapter 12

# Linear Regression and Correlation

### 12.1 95% Critical Values of the Sample Correlation Coefficient Table<sup>1</sup>

Degrees of Freedom: $n - 2$	Critical Values: (+ and -)
1	0.997
2	0.950
3	0.878
4	0.811
5	0.754
6	0.707
7	0.666
8	0.632
9	0.602
10	0.576
11	0.555
12	0.532
13	0.514
14	0.497
15	0.482
16	0.468

*continued on next page*

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m17098/1.5/>>.

17	0.456
18	0.444
19	0.433
20	0.423
21	0.413
22	0.404
23	0.396
24	0.388
25	0.381
26	0.374
27	0.367
28	0.361
29	0.355
30	0.349
40	0.304
50	0.273
60	0.250
70	0.232
80	0.217
90	0.205
100 and over	0.195

Table 12.1

## 12.2 Summary<sup>2</sup>

**Bivariate Data:** Each data point has two values. The form is  $(x, y)$ .

**Line of Best Fit or Least Squares Line (LSL):**  $\hat{y} = a + bx$

$x$  = independent variable;  $y$  = dependent variable

**Residual:** Actual  $y$  value – predicted  $y$  value =  $y - \hat{y}$

**Correlation Coefficient  $r$ :**

1. Used to determine whether a line of best fit is good for prediction.
2. Between -1 and 1 inclusive. The closer  $r$  is to 1 or -1, the closer the original points are to a straight line.
3. If  $r$  is negative, the slope is negative. If  $r$  is positive, the slope is positive.
4. If  $r = 0$ , then the line is horizontal.

---

<sup>2</sup>This content is available online at <<http://cnx.org/content/m17081/1.4/>>.

**Sum of Squared Errors (SSE):** The smaller the SSE, the better the original set of points fits the line of best fit.

**Outlier:** A point that does not seem to fit the rest of the data.

## 12.3 Homework<sup>3</sup>

### Exercise 12.1

(*Solution on p. 204.*)

For each situation below, state the independent variable and the dependent variable.

- A study is done to determine if elderly drivers are involved in more motor vehicle fatalities than all other drivers. The number of fatalities per 100,000 drivers is compared to the age of drivers.
- A study is done to determine if the weekly grocery bill changes based on the number of family members.
- Insurance companies base life insurance premiums partially on the age of the applicant.
- Utility bills vary according to power consumption.
- A study is done to determine if a higher education reduces the crime rate in a population.

### Exercise 12.2

In 1990 the number of driver deaths per 100,000 for the different age groups was as follows (Source: *The National Highway Traffic Safety Administration's National Center for Statistics and Analysis*):

Age	Number of Driver Deaths per 100,000
15-24	28
25-39	15
40-69	10
70-79	15
80+	25

Table 12.2

- For each age group, pick the midpoint of the interval for the x value. (For the 80+ group, use 85.)
- Using “ages” as the independent variable and “Number of driver deaths per 100,000” as the dependent variable, make a scatter plot of the data.
- Calculate the least squares (best-fit) line. Put the equation in the form of:  $y = a + bx$
- Find the correlation coefficient. Is it significant?
- Pick two ages and find the estimated fatality rates.
- Use the two points in (e) to plot the least squares line on your graph from (b).
- Based on the above data, is there a linear relationship between age of a driver and driver fatality rate?
- What is the slope of the least squares (best-fit) line? Interpret the slope.

<sup>3</sup>This content is available online at <<http://cnx.org/content/m33266/1.1/>>.

**Exercise 12.3**

(Solution on p. 204.)

The average number of people in a family that received welfare for various years is given below.  
 (Source: House Ways and Means Committee, Health and Human Services Department)

Year	Welfare family size
1969	4.0
1973	3.6
1975	3.2
1979	3.0
1983	3.0
1988	3.0
1991	2.9

**Table 12.3**

- a. Using "year" as the independent variable and "welfare family size" as the dependent variable, make a scatter plot of the data.
- b. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- c. Find the correlation coefficient. Is it significant?
- d. Pick two years between 1969 and 1991 and find the estimated welfare family sizes.
- e. Use the two points in (d) to plot the least squares line on your graph from (b).
- f. Based on the above data, is there a linear relationship between the year and the average number of people in a welfare family?
- g. Using the least squares line, estimate the welfare family sizes for 1960 and 1995. Does the least squares line give an accurate estimate for those years? Explain why or why not.
- h. Are there any outliers in the above data?
- i. What is the estimated average welfare family size for 1986? Does the least squares line give an accurate estimate for that year? Explain why or why not.
- j. What is the slope of the least squares (best-fit) line? Interpret the slope.

**Exercise 12.4**

Use the AIDS data from the practice for this section<sup>4</sup>, but this time use the columns "year #" and "# new AIDS deaths in U.S." Answer all of the questions from the practice again, using the new columns.

**Exercise 12.5**

(Solution on p. 204.)

The height (sidewalk to roof) of notable tall buildings in America is compared to the number of stories of the building (beginning at street level). (Source: Microsoft Bookshelf)

---

<sup>4</sup>"Linear Regression and Correlation: Practice": Section Given <<http://cnx.org/content/m17088/latest/#element-476>>

Height (in feet)	Stories
1050	57
428	28
362	26
529	40
790	60
401	22
380	38
1454	110
1127	100
700	46

Table 12.4

- Using "stories" as the independent variable and "height" as the dependent variable, make a scatter plot of the data.
- Does it appear from inspection that there is a relationship between the variables?
- Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- Find the correlation coefficient. Is it significant?
- Find the estimated heights for 32 stories and for 94 stories.
- Use the two points in (e) to plot the least squares line on your graph from (b).
- Based on the above data, is there a linear relationship between the number of stories in tall buildings and the height of the buildings?
- Are there any outliers in the above data? If so, which point(s)?
- What is the estimated height of a building with 6 stories? Does the least squares line give an accurate estimate of height? Explain why or why not.
- Based on the least squares line, adding an extra story adds about how many feet to a building?
- What is the slope of the least squares (best-fit) line? Interpret the slope.

### Exercise 12.6

Below is the life expectancy for an individual born in the United States in certain years. (Source: *National Center for Health Statistics*)

Year of Birth	Life Expectancy
1930	59.7
1940	62.9
1950	70.2
1965	69.7
1973	71.4
1982	74.5
1987	75
1992	75.7

**Table 12.5**

- a. Decide which variable should be the independent variable and which should be the dependent variable.
- b. Draw a scatter plot of the ordered pairs.
- c. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- d. Find the correlation coefficient. Is it significant?
- e. Find the estimated life expectancy for an individual born in 1950 and for one born in 1982.
- f. Why aren't the answers to part (e) the values on the above chart that correspond to those years?
- g. Use the two points in (e) to plot the least squares line on your graph from (b).
- h. Based on the above data, is there a linear relationship between the year of birth and life expectancy?
- i. Are there any outliers in the above data?
- j. Using the least squares line, find the estimated life expectancy for an individual born in 1850. Does the least squares line give an accurate estimate for that year? Explain why or why not.
- k. What is the slope of the least squares (best-fit) line? Interpret the slope.

**Exercise 12.7**

(Solution on p. 204.)

The percent of female wage and salary workers who are paid hourly rates is given below for the years 1979 - 1992. (Source: Bureau of Labor Statistics, U.S. Dept. of Labor)

Year	Percent of workers paid hourly rates
1979	61.2
1980	60.7
1981	61.3
1982	61.3
1983	61.8
1984	61.7
1985	61.8
1986	62.0
1987	62.7
1990	62.8
1992	62.9

**Table 12.6**

- a. Using "year" as the independent variable and "percent" as the dependent variable, make a scatter plot of the data.
- b. Does it appear from inspection that there is a relationship between the variables? Why or why not?
- c. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- d. Find the correlation coefficient. Is it significant?
- e. Find the estimated percents for 1991 and 1988.
- f. Use the two points in (e) to plot the least squares line on your graph from (b).

- g. Based on the above data, is there a linear relationship between the year and the percent of female wage and salary earners who are paid hourly rates?
- h. Are there any outliers in the above data?
- i. What is the estimated percent for the year 2050? Does the least squares line give an accurate estimate for that year? Explain why or why not?
- j. What is the slope of the least squares (best-fit) line? Interpret the slope.

**Exercise 12.8**

The maximum discount value of the Entertainment® card for the “Fine Dining” section, Edition 10, for various pages is given below.

Page number	Maximum value (\$)
4	16
14	19
25	15
32	17
43	19
57	15
72	16
85	15
90	17

Table 12.7

- a. Decide which variable should be the independent variable and which should be the dependent variable.
- b. Draw a scatter plot of the ordered pairs.
- c. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- d. Find the correlation coefficient. Is it significant?
- e. Find the estimated maximum values for the restaurants on page 10 and on page 70.
- f. Use the two points in (e) to plot the least squares line on your graph from (b).
- g. Does it appear that the restaurants giving the maximum value are placed in the beginning of the “Fine Dining” section? How did you arrive at your answer?
- h. Suppose that there were 200 pages of restaurants. What do you estimate to be the maximum value for a restaurant listed on page 200?
- i. Is the least squares line valid for page 200? Why or why not?
- j. What is the slope of the least squares (best-fit) line? Interpret the slope.

**The next two questions refer to the following data:** The cost of a leading liquid laundry detergent in different sizes is given below.

Size (ounces)	Cost (\$)	Cost per ounce
16	3.99	
32	4.99	
64	5.99	
200	10.99	

**Table 12.8****Exercise 12.9***(Solution on p. 204.)*

- a. Using "size" as the independent variable and "cost" as the dependent variable, make a scatter plot.
- b. Does it appear from inspection that there is a relationship between the variables? Why or why not?
- c. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- d. Find the correlation coefficient. Is it significant?
- e. If the laundry detergent were sold in a 40 ounce size, find the estimated cost.
- f. If the laundry detergent were sold in a 90 ounce size, find the estimated cost.
- g. Use the two points in (e) and (f) to plot the least squares line on your graph from (a).
- h. Does it appear that a line is the best way to fit the data? Why or why not?
- i. Are there any outliers in the above data?
- j. Is the least squares line valid for predicting what a 300 ounce size of the laundry detergent would cost? Why or why not?
- k. What is the slope of the least squares (best-fit) line? Interpret the slope.

**Exercise 12.10**

- a. Complete the above table for the cost per ounce of the different sizes.
- b. Using "Size" as the independent variable and "Cost per ounce" as the dependent variable, make a scatter plot of the data.
- c. Does it appear from inspection that there is a relationship between the variables? Why or why not?
- d. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- e. Find the correlation coefficient. Is it significant?
- f. If the laundry detergent were sold in a 40 ounce size, find the estimated cost per ounce.
- g. If the laundry detergent were sold in a 90 ounce size, find the estimated cost per ounce.
- h. Use the two points in (f) and (g) to plot the least squares line on your graph from (b).
- i. Does it appear that a line is the best way to fit the data? Why or why not?
- j. Are there any outliers in the above data?
- k. Is the least squares line valid for predicting what a 300 ounce size of the laundry detergent would cost per ounce? Why or why not?
- l. What is the slope of the least squares (best-fit) line? Interpret the slope.

**Exercise 12.11***(Solution on p. 204.)*

According to flyer by a Prudential Insurance Company representative, the costs of approximate probate fees and taxes for selected net taxable estates are as follows:

Net Taxable Estate (\$)	Approximate Probate Fees and Taxes (\$)
600,000	30,000
750,000	92,500
1,000,000	203,000
1,500,000	438,000
2,000,000	688,000
2,500,000	1,037,000
3,000,000	1,350,000

**Table 12.9**

- a. Decide which variable should be the independent variable and which should be the dependent variable.
- b. Make a scatter plot of the data.
- c. Does it appear from inspection that there is a relationship between the variables? Why or why not?
- d. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- e. Find the correlation coefficient. Is it significant?
- f. Find the estimated total cost for a net taxable estate of \$1,000,000. Find the cost for \$2,500,000.
- g. Use the two points in (f) to plot the least squares line on your graph from (b).
- h. Does it appear that a line is the best way to fit the data? Why or why not?
- i. Are there any outliers in the above data?
- j. Based on the above, what would be the probate fees and taxes for an estate that does not have any assets?
- k. What is the slope of the least squares (best-fit) line? Interpret the slope.

**Exercise 12.12**

The following are advertised sale prices of color televisions at Anderson's.

Size (inches)	Sale Price (\$)
9	147
20	197
27	297
31	447
35	1177
40	2177
60	2497

**Table 12.10**

- a. Decide which variable should be the independent variable and which should be the dependent variable.
- b. Make a scatter plot of the data.
- c. Does it appear from inspection that there is a relationship between the variables? Why or why not?
- d. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- e. Find the correlation coefficient. Is it significant?
- f. Find the estimated sale price for a 32 inch television. Find the cost for a 50 inch television.
- g. Use the two points in (f) to plot the least squares line on your graph from (b).
- h. Does it appear that a line is the best way to fit the data? Why or why not?
- i. Are there any outliers in the above data?
- j. What is the slope of the least squares (best-fit) line? Interpret the slope.

**Exercise 12.13**

(*Solution on p. 205.*)

Below are the average heights for American boys. (Source: *Physician's Handbook, 1990*)

Age (years)	Height (cm)
birth	50.8
2	83.8
3	91.4
5	106.6
7	119.3
10	137.1
14	157.5

**Table 12.11**

- a. Decide which variable should be the independent variable and which should be the dependent variable.
- b. Make a scatter plot of the data.
- c. Does it appear from inspection that there is a relationship between the variables? Why or why not?
- d. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- e. Find the correlation coefficient. Is it significant?
- f. Find the estimated average height for a one year-old. Find the estimated average height for an eleven year-old.
- g. Use the two points in (f) to plot the least squares line on your graph from (b).
- h. Does it appear that a line is the best way to fit the data? Why or why not?
- i. Are there any outliers in the above data?
- j. Use the least squares line to estimate the average height for a sixty-two year-old man. Do you think that your answer is reasonable? Why or why not?
- k. What is the slope of the least squares (best-fit) line? Interpret the slope.

**Exercise 12.14**

The following chart gives the gold medal times for every other Summer Olympics for the women's 100 meter freestyle (swimming).

Year	Time (seconds)
1912	82.2
1924	72.4
1932	66.8
1952	66.8
1960	61.2
1968	60.0
1976	55.65
1984	55.92
1992	54.64

**Table 12.12**

- a. Decide which variable should be the independent variable and which should be the dependent variable.
- b. Make a scatter plot of the data.
- c. Does it appear from inspection that there is a relationship between the variables? Why or why not?
- d. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- e. Find the correlation coefficient. Is the decrease in times significant?
- f. Find the estimated gold medal time for 1932. Find the estimated time for 1984.
- g. Why are the answers from (f) different from the chart values?
- h. Use the two points in (f) to plot the least squares line on your graph from (b).
- i. Does it appear that a line is the best way to fit the data? Why or why not?
- j. Use the least squares line to estimate the gold medal time for the next Summer Olympics. Do you think that your answer is reasonable? Why or why not?

The next three questions use the following state information.

State	# letters in name	Year entered the Union	Rank for entering the Union	Area (square miles)
Alabama	7	1819	22	52,423
Colorado		1876	38	104,100
Hawaii		1959	50	10,932
Iowa		1846	29	56,276
Maryland		1788	7	12,407
Missouri		1821	24	69,709
New Jersey		1787	3	8,722
Ohio		1803	17	44,828
South Carolina	13	1788	8	32,008
Utah		1896	45	84,904
Wisconsin		1848	30	65,499

Table 12.13

**Exercise 12.15**

(*Solution on p. 205.*)

We are interested in whether or not the number of letters in a state name depends upon the year the state entered the Union.

- a. Decide which variable should be the independent variable and which should be the dependent variable.
- b. Make a scatter plot of the data.
- c. Does it appear from inspection that there is a relationship between the variables? Why or why not?
- d. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- e. Find the correlation coefficient. What does it imply about the significance of the relationship?
- f. Find the estimated number of letters (to the nearest integer) a state would have if it entered the Union in 1900. Find the estimated number of letters a state would have if it entered the Union in 1940.

- g. Use the two points in (f) to plot the least squares line on your graph from (b).
- h. Does it appear that a line is the best way to fit the data? Why or why not?
- i. Use the least squares line to estimate the number of letters a new state that enters the Union this year would have. Can the least squares line be used to predict it? Why or why not?

**Exercise 12.16**

We are interested in whether there is a relationship between the ranking of a state and the area of the state.

- a. Let rank be the independent variable and area be the dependent variable.
- b. What do you think the scatter plot will look like? Make a scatter plot of the data.
- c. Does it appear from inspection that there is a relationship between the variables? Why or why not?
- d. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- e. Find the correlation coefficient. What does it imply about the significance of the relationship?
- f. Find the estimated areas for Alabama and for Colorado. Are they close to the actual areas?
- g. Use the two points in (f) to plot the least squares line on your graph from (b).
- h. Does it appear that a line is the best way to fit the data? Why or why not?
- i. Are there any outliers?
- j. Use the least squares line to estimate the area of a new state that enters the Union. Can the least squares line be used to predict it? Why or why not?
- k. Delete "Hawaii" and substitute "Alaska" for it. Alaska is the fortieth state with an area of 656,424 square miles.
- l. Calculate the new least squares line.
- m. Find the estimated area for Alabama. Is it closer to the actual area with this new least squares line or with the previous one that included Hawaii? Why do you think that's the case?
- n. Do you think that, in general, newer states are larger than the original states?

**Exercise 12.17**

(Solution on p. 205.)

We are interested in whether there is a relationship between the rank of a state and the year it entered the Union.

- a. Let year be the independent variable and rank be the dependent variable.
- b. What do you think the scatter plot will look like? Make a scatter plot of the data.
- c. Why must the relationship be positive between the variables?
- d. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- e. Find the correlation coefficient. What does it imply about the significance of the relationship?
- f. Let's say a fifty-first state entered the union. Based upon the least squares line, when should that have occurred?
- g. Using the least squares line, how many states do we currently have?
- h. Why isn't the least squares line a good estimator for this year?

**Exercise 12.18**

Below are the percents of the U.S. labor force (excluding self-employed and unemployed) that are members of a union. We are interested in whether the decrease is significant. (Source: *Bureau of Labor Statistics, U.S. Dept. of Labor*)

Year	Percent
1945	35.5
1950	31.5
1960	31.4
1970	27.3
1980	21.9
1986	17.5
1993	15.8

**Table 12.14**

- a. Let year be the independent variable and percent be the dependent variable.
- b. What do you think the scatter plot will look like? Make a scatter plot of the data.
- c. Why will the relationship between the variables be negative?
- d. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- e. Find the correlation coefficient. What does it imply about the significance of the relationship?
- f. Based on your answer to (e), do you think that the relationship can be said to be decreasing?
- g. If the trend continues, when will there no longer be any union members? Do you think that will happen?

**The next two questions refer to the following information:** The data below reflects the 1991-92 Reunion Class Giving. (Source: *SUNY Albany alumni magazine*)

Class Year	Average Gift	Total Giving
1922	41.67	125
1927	60.75	1,215
1932	83.82	3,772
1937	87.84	5,710
1947	88.27	6,003
1952	76.14	5,254
1957	52.29	4,393
1962	57.80	4,451
1972	42.68	18,093
1976	49.39	22,473
1981	46.87	20,997
1986	37.03	12,590

**Table 12.15****Exercise 12.19**

(Solution on p. 205.)

We will use the columns “class year” and “total giving” for all questions, unless otherwise stated.

- a. What do you think the scatter plot will look like? Make a scatter plot of the data.
- b. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- c. Find the correlation coefficient. What does it imply about the significance of the relationship?
- d. For the class of 1930, predict the total class gift.
- e. For the class of 1964, predict the total class gift.
- f. For the class of 1850, predict the total class gift. Why doesn't this value make any sense?

**Exercise 12.20**

We will use the columns "class year" and "average gift" for all questions, unless otherwise stated.

- a. What do you think the scatter plot will look like? Make a scatter plot of the data.
- b. Calculate the least squares line. Put the equation in the form of:  $\hat{y} = a + bx$
- c. Find the correlation coefficient. What does it imply about the significance of the relationship?
- d. For the class of 1930, predict the average class gift.
- e. For the class of 1964, predict the average class gift.
- f. For the class of 2010, predict the average class gift. Why doesn't this value make any sense?

**12.3.1 Try these multiple choice questions****Exercise 12.21**

(*Solution on p. 205.*)

A correlation coefficient of -0.95 means there is a \_\_\_\_\_ between the two variables.

- A. Strong positive correlation
- B. Weak negative correlation
- C. Strong negative correlation
- D. No Correlation

**Exercise 12.22**

(*Solution on p. 205.*)

According to the data reported by the New York State Department of Health regarding West Nile Virus for the years 2000-2004, the least squares line equation for the number of reported dead birds

$(x)$  versus the number of human West Nile virus cases  $(y)$  is  $\hat{y} = -10.2638 + 0.0491x$ . If the number of dead birds reported in a year is 732, how many human cases of West Nile virus can be expected?

- A. 25.7
- B. 46.2
- C. -25.7
- D. 7513

**The next three questions refer to the following data:** (showing the number of hurricanes by category to directly strike the mainland U.S. each decade) obtained from [www.nhc.noaa.gov/gifs/table6.gif](http://www.nhc.noaa.gov/gifs/table6.gif)<sup>5</sup> A major hurricane is one with a strength rating of 3, 4 or 5.

---

<sup>5</sup><http://www.nhc.noaa.gov/gifs/table6.gif>

Decade	Total Number of Hurricanes	Number of Major Hurricanes
1941-1950	24	10
1951-1960	17	8
1961-1970	14	6
1971-1980	12	4
1981-1990	15	5
1991-2000	14	5
2001 – 2004	9	3

**Table 12.16****Exercise 12.23***(Solution on p. 205.)*

Using only completed decades (1941 – 2000), calculate the least squares line for the number of major hurricanes expected based upon the total number of hurricanes.

- A.  $\hat{y} = -1.67x + 0.5$
- B.  $\hat{y} = 0.5x - 1.67$
- C.  $\hat{y} = 0.94x - 1.67$
- D.  $\hat{y} = -2x + 1$

**Exercise 12.24***(Solution on p. 205.)*

The correlation coefficient is 0.942. Is this considered significant? Why or why not?

- A. No, because 0.942 is greater than the critical value of 0.707
- B. Yes, because 0.942 is greater than the critical value of 0.707
- C. No, because 0.942 is greater than the critical value of 0.811
- D. Yes, because 0.942 is greater than the critical value of 0.811

**Exercise 12.25***(Solution on p. 205.)*

The data for 2001-2004 show 9 hurricanes have hit the mainland United States. The line of best fit predicts 2.83 major hurricanes to hit mainland U.S. Can the least squares line be used to make this prediction?

- A. No, because 9 lies outside the independent variable values
- B. Yes, because, in fact, there have been 3 major hurricanes this decade
- C. No, because 2.83 lies outside the dependent variable values
- D. Yes, because how else could we predict what is going to happen this decade.

**Exercise 12.26***(Solution on p. 205.)*

We are interested in exploring the relationship between the weight of a vehicle and its fuel efficiency (gasoline mileage). The data in the table show the weights, in pounds, and fuel efficiency, measured in miles per gallon, for a sample of 12 vehicles.

Weight	Fuel Efficiency
2715	24
2570	28
2610	29
2750	38
3000	25
3410	22
3640	20
3700	26
3880	21
3900	18
4060	18
4710	15

Table 12.17

- a. Graph a scatterplot of the data.
- b. Find the correlation coefficient and determine if it is significant.
- c. Find the equation of the best fit line.
- d. Write the sentence that interprets the meaning of the slope of the line in the context of the data.
- e. What percent of the variation in fuel efficiency is explained by the variation in the weight of the vehicles, using the regression line? (State your answer in a complete sentence in the context of the data.)
- f. Accurately graph the best fit line on your scatterplot.
- g. For the vehicle that weights 3000 pounds, find the residual ( $y - \hat{y}$ ). Does the value predicted by the line underestimate or overestimate the observed data value?
- h. Identify any outliers, using either the graphical or numerical procedure demonstrated in the textbook.
- i. The outlier is a hybrid car that runs on gasoline and electric technology, but all other vehicles in the sample have engines that use gasoline only. Explain why it would be appropriate to remove the outlier from the data in this situation. Remove the outlier from the sample data. Find the new correlation coefficient, coefficient of determination, and best fit line.
- j. Compare the correlation coefficients and coefficients of determination before and after removing the outlier, and explain in complete sentences what these numbers indicate about how the model has changed.

**Exercise 12.27**

(Solution on p. 206.)

The four data sets below were created by statistician Francis Anscomb. They show why it is important to examine the scatterplots for your data, in addition to finding the correlation coefficient, in order to evaluate the appropriateness of fitting a linear model.

Set 1		Set 2		Set 3		Set 4	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

**Table 12.18**

- a. For each data set, find the least squares regression line and the correlation coefficient. What did you discover about the lines and values of  $r$ ?

For each data set, create a scatter plot and graph the least squares regression line. Use the graphs to answer the following questions:

- b. For which data set does it appear that a curve would be a more appropriate model than a line?
- c. Which data set has an **influential point** (point close to or on the line that greatly influences the best fit line)?
- d. Which data set has an **outlier** (obviously visible on the scatter plot with best fit line graphed)?
- e. Which data set appears to be the most appropriate to model using the least squares regression line?

## Solutions to Exercises in Chapter 12

### Solution to Exercise 12.1 (p. 189)

- a. Independent: Age; Dependent: Fatalities
- d. Independent: Power Consumption; Dependent: Utility

### Solution to Exercise 12.3 (p. 190)

- b.  $\hat{y} = 88.7206 - 0.0432x$
- c. -0.8533, Yes
- g. No
- h. No.
- i. 2.97, Yes
- j. slope = -0.0432. As the year increases by one, the welfare family size decreases by 0.0432 people.

### Solution to Exercise 12.5 (p. 190)

- b. Yes
- c.  $\hat{y} = 102.4287 + 11.7585x$
- d. 0.9436; yes
- e. 478.70 feet; 1207.73 feet
- g. Yes
- h. Yes; (57, 1050)
- i. 172.98; No
- j. 11.7585 feet
- k. slope = 11.7585. As the number of stories increases by one, the height of the building increases by 11.7585 feet.

### Solution to Exercise 12.7 (p. 192)

- b. Yes
- c.  $\hat{y} = -266.8863 + 0.1656x$
- d. 0.9448; Yes
- e. 62.9206; 62.4237
- h. No
- i. 72.639; No
- j. slope = 0.1656. As the year increases by one, the percent of workers paid hourly rates increases by 0.1565.

### Solution to Exercise 12.9 (p. 194)

- b. Yes
- c.  $\hat{y} = 3.5984 + 0.0371x$
- d. 0.9986; Yes
- e. \$5.08
- f. \$6.93
- i. No
- j. Not valid
- k. slope = 0.0371. As the number of ounces increases by one, the cost of the liquid detergent increases by \$0.0371 (or about 4 cents).

### Solution to Exercise 12.11 (p. 194)

- c. Yes

- d.  $\hat{y} = -337,424.6478 + 0.5463x$   
 e. 0.9964; Yes  
 f. \$208,872.49; \$1,028,318.20  
 h. Yes  
 i. No  
 k. slope = 0.5463. As the net taxable estate increases by one dollar, the approximate probate fees and taxes increases by 0.5463 dollars (about 55 cents).

**Solution to Exercise 12.13 (p. 195)**

- c. Yes  
 $\hat{y}$   
 d.  $\hat{y} = 65.0876 + 7.0948x$   
 e. 0.9761; yes  
 f. 72.2 cm; 143.13 cm  
 h. Yes  
 i. No  
 j. 505.0 cm; No  
 k. slope = 7.0948. As the age of an American boy increases by one year, the average height increases by 7.0948 cm.

**Solution to Exercise 12.15 (p. 197)**

- c. No  
 $\hat{y}$   
 d.  $\hat{y} = 47.03 - 0.216x$   
 e. -0.4280  
 f. 6; 5

**Solution to Exercise 12.17 (p. 198)**

- $\hat{y}$   
 d.  $\hat{y} = -480.5845 + 0.2748x$   
 e. 0.9553  
 f. 1934

**Solution to Exercise 12.19 (p. 199)**

- b.  $\hat{y} = -569,770.2796 + 296.0351x$   
 c. 0.8302  
 d. \$1577.48  
 e. \$11,642.68  
 f. -\$22,105.33

**Solution to Exercise 12.21 (p. 200)**

C

**Solution to Exercise 12.22 (p. 200)**

A

**Solution to Exercise 12.23 (p. 201)**

A

**Solution to Exercise 12.24 (p. 201)**

D

**Solution to Exercise 12.25 (p. 201)**

A

**Solution to Exercise 12.26 (p. 201)**

- b.  $r = -0.8$ , significant
- c.  $\hat{y} = 48.4 - 0.00725x$
- d. For every one pound increase in weight, the fuel efficiency decreases by 0.00725 miles per gallon. (For every one thousand pound increase in weight, the fuel efficiency decreases by 7.25 miles per gallon.)
- e. 64% of the variation in fuel efficiency is explained by the variation in weight using the regression line.
- g.  $\hat{y} = 48.4 - 0.00725(3000) = 26.65$  mpg.  $y - \hat{y} = 25 - 26.65 = -1.65$ . Because  $\hat{y} = 26.5$  is greater than  $y = 25$ , the line overestimates the observed fuel efficiency.
- h. (2750, 38) is the outlier. Be sure you know how to justify it using the requested graphical or numerical methods, not just by guessing.
- i.  $\hat{y} = 42.4 - 0.00578x$
- j. Without outlier,  $r = -0.885$ ,  $R^2 = 0.76$ ; with outlier,  $r = -0.8$ ,  $R^2 = 0.64$ . The new linear model is a better fit, after the outlier is removed from the data, because the new correlation coefficient is farther from 0 and the new coefficient of determination is larger.

**Solution to Exercise 12.27 (p. 202)**

a. All four data sets have the same correlation coefficient  $r=0.816$  and the same least squares regression line  $\hat{y}=3+0.5x$

b. Set 2 ; c. Set 4 ; d. Set 3 ; e. Set 1

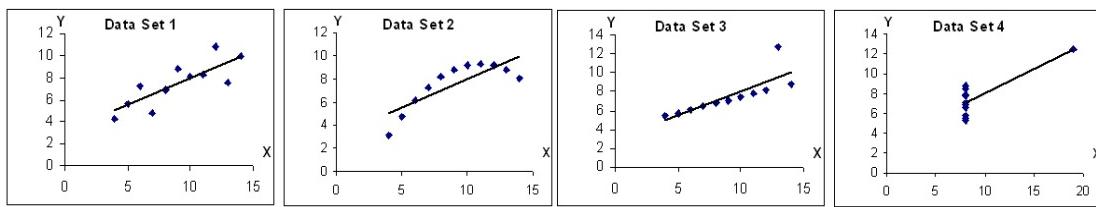


Figure 12.1

# Chapter 13

## F Distribution and ANOVA

### 13.1 Summary<sup>1</sup>

- An ANOVA hypothesis test determines if several population means are equal. The distribution for the test is the F distribution with 2 different degrees of freedom.

**Assumptions:**

- a. Each population from which a sample is taken is assumed to be normal.
- b. Each sample is randomly selected and independent.
- c. The populations are assumed to have equal standard deviations (or variances)

- A Test of Two Variances hypothesis test determines if two variances are the same. The distribution for the hypothesis test is the F distribution with 2 different degrees of freedom.

**Assumptions:**

- a. The populations from which the two samples are drawn are normally distributed.
- b. The two populations are independent of each other.

### 13.2 Homework<sup>2</sup>

DIRECTIONS: Use a solution sheet to conduct the following hypothesis tests. The solution sheet can be found in the Table of Contents 14. Appendix.

**Exercise 13.1**

*(Solution on p. 213.)*

Three students, Linda, Tuan, and Javier, are given 5 laboratory rats each for a nutritional experiment. Each rat's weight is recorded in grams. Linda feeds her rats Formula A, Tuan feeds his rats Formula B, and Javier feeds his rats Formula C. At the end of a specified time period, each rat is weighed again and the net gain in grams is recorded. Using a significance level of 10%, test the hypothesis that the three formulas produce the same average weight gain.

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m17072/1.3/>>.

<sup>2</sup>This content is available online at <<http://cnx.org/content/m17063/1.9/>>.

**Weights of Student Lab Rats**

Linda's rats	Tuan's rats	Javier's rats
43.5	47.0	51.2
39.4	40.5	40.9
41.3	38.9	37.9
46.0	46.3	45.0
38.2	44.2	48.6

**Table 13.1****Exercise 13.2**

A grassroots group opposed to a proposed increase in the gas tax claimed that the increase would hurt working-class people the most, since they commute the farthest to work. Suppose that the group randomly surveyed 24 individuals and asked them their daily one-way commuting mileage. The results are below:

working-class	professional (middle incomes)	professional (wealthy)
17.8	16.5	8.5
26.7	17.4	6.3
49.4	22.0	4.6
9.4	7.4	12.6
65.4	9.4	11.0
47.1	2.1	28.6
19.5	6.4	15.4
51.2	13.9	9.3

**Table 13.2****Exercise 13.3***(Solution on p. 213.)*

Refer to Exercise 13.8.1. Determine whether or not the variance in weight gain is statistically the same among Javier's and Linda's rats.

**Exercise 13.4**

Refer to Exercise 13.8.2 above (Exercise 13.2). Determine whether or not the variance in mileage driven is statistically the same among the working class and professional (middle income) groups.

**For the next two problems, refer to the data from Terri Vogel's Log Book [link pending].**

**Exercise 13.5***(Solution on p. 213.)*

Examine the 7 practice laps. Determine whether the average lap time is statistically the same for the 7 practice laps, or if there is at least one lap that has a different average time from the others.

**Exercise 13.6**

Examine practice laps 3 and 4. Determine whether or not the variance in lap time is statistically the same for those practice laps.

**For the next four problems, refer to the following data.**

The following table lists the number of pages in four different types of magazines.

home decorating	news	health	computer
172	87	82	104
286	94	153	136
163	123	87	98
205	106	103	207
197	101	96	146

**Table 13.3****Exercise 13.7***(Solution on p. 213.)*

Using a significance level of 5%, test the hypothesis that the four magazine types have the same average length.

**Exercise 13.8**

Eliminate one magazine type that you now feel has an average length different than the others. Redo the hypothesis test, testing that the remaining three averages are statistically the same. Use a new solution sheet. Based on this test, are the average lengths for the remaining three magazines statistically the same?

**Exercise 13.9**

Which two magazine types do you think have the same variance in length?

**Exercise 13.10**

Which two magazine types do you think have different variances in length?

### 13.3 Review Questions<sup>3</sup>

**The next two questions refer to the following situation:**

Suppose that the probability of a drought in any independent year is 20%. Out of those years in which a drought occurs, the probability of water rationing is 10%. However, in any year, the probability of water rationing is 5%.

**Exercise 13.11: REVIEW QUESTION 1***(Solution on p. 213.)*

What is the probability of both a drought **and** water rationing occurring?

**Exercise 13.12: REVIEW QUESTION 2***(Solution on p. 213.)*

Out of the years with water rationing, find the probability that there is a drought.

**The next three questions refer to the following survey:**

#### Favorite Type of Pie by Gender

	apple	pumpkin	pecan
female	40	10	30
male	20	30	10

**Table 13.4**


---

<sup>3</sup>This content is available online at <<http://cnx.org/content/m19022/1.1/>>.

**Exercise 13.13: REVIEW QUESTION 3**

(Solution on p. 213.)

Suppose that one individual is randomly chosen. Find the probability that the person's favorite pie is apple or the person is male.

**Exercise 13.14: REVIEW QUESTION 4**

(Solution on p. 213.)

Suppose that one male is randomly chosen. Find the probability his favorite pie is pecan.

**Exercise 13.15: REVIEW QUESTION 5**

(Solution on p. 213.)

Conduct a hypothesis test to determine if favorite pie type and gender are independent.

**The next two questions refer to the following situation:**

Let's say that the probability that an adult watches the news at least once per week is 0.60.

**Exercise 13.16: REVIEW QUESTION 6**

(Solution on p. 213.)

We randomly survey 14 people. On average, how many people do we expect to watch the news at least once per week?

**Exercise 13.17: REVIEW QUESTION 7**

(Solution on p. 213.)

We randomly survey 14 people. Of interest is the number that watch the news at least once per week. State the distribution of  $X$ .  $X \sim$

**Exercise 13.18: REVIEW QUESTION 8**

(Solution on p. 213.)

The following histogram is most likely to be a result of sampling from which distribution?

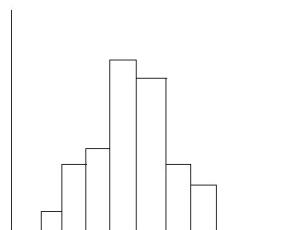


Figure 13.1

- A. Chi-Square
- B. Geometric
- C. Uniform
- D. Binomial

**Exercise 13.19**

REVIEW QUESTION 9 Removed

**The next three questions refer to the following situation:**

The amount of money a customer spends in one trip to the supermarket is known to have an exponential distribution. Suppose the average amount of money a customer spends in one trip to the supermarket is \$72.

**Exercise 13.20: REVIEW QUESTION 10**

(Solution on p. 213.)

Find the probability that one customer spends less than \$72 in one trip to the supermarket?

**Exercise 13.21: REVIEW QUESTION 11**

(Solution on p. 213.)

Suppose 5 customers pool their money. (They are poor college students.) How much money altogether would you expect the 5 customers to spend in one trip to the supermarket (in dollars)?

**Exercise 13.22: REVIEW QUESTION 12**

(Solution on p. 213.)

State the distribution to use if you want to find the probability that the **average** amount spent by 5 customers in one trip to the supermarket is less than \$60.

**Exercise 13.23: REVIEW QUESTION 13**

(Solution on p. 213.)

A math exam was given to all the fifth grade children attending Country School. Two random samples of scores were taken. The null hypothesis is that the average math scores for boys and girls in fifth grade are the same. Conduct a hypothesis test.

	$n$	$\bar{x}$	$s^2$
Boys	55	82	29
Girls	60	86	46

**Table 13.5****Exercise 13.24: REVIEW QUESTION 14**

(Solution on p. 214.)

In a survey of 80 males, 55 had played an organized sport growing up. Of the 70 females surveyed, 25 had played an organized sport growing up. We are interested in whether the proportion for males is higher than the proportion for females. Conduct a hypothesis test.

**Exercise 13.25: REVIEW QUESTION 15**

(Solution on p. 214.)

Which of the following is preferable when designing a hypothesis test?

- A. Maximize  $\alpha$  and minimize  $\beta$
- B. Minimize  $\alpha$  and maximize  $\beta$
- C. Maximize  $\alpha$  and  $\beta$
- D. Minimize  $\alpha$  and  $\beta$

The next three questions refer to the following situation:

120 people were surveyed as to their favorite beverage (non-alcoholic). The results are below.

**Preferred Beverage by Age**

	0 – 9	10 – 19	20 – 29	30 +	Totals
Milk	14	10	6	0	30
Soda	3	8	26	15	52
Juice	7	12	12	7	38
Totals	24	30	44	22	120

**Table 13.6****Exercise 13.26: REVIEW QUESTION 16**

(Solution on p. 214.)

Are the events of **milk** and **30+**:

- a. Independent events? Justify your answer.
- b. Mutually exclusive events? Justify your answer.

**Exercise 13.27: REVIEW QUESTION 17***(Solution on p. 214.)*

Suppose that one person is randomly chosen. Find the probability that person is 10 – 19 given that he/she prefers juice.

**Exercise 13.28: REVIEW QUESTION 18***(Solution on p. 214.)*

Are Preferred Beverage and Age independent events? Conduct a hypothesis test.

**Exercise 13.29: REVIEW QUESTION 19***(Solution on p. 214.)*

Given the following histogram, which distribution is the data most likely to come from?

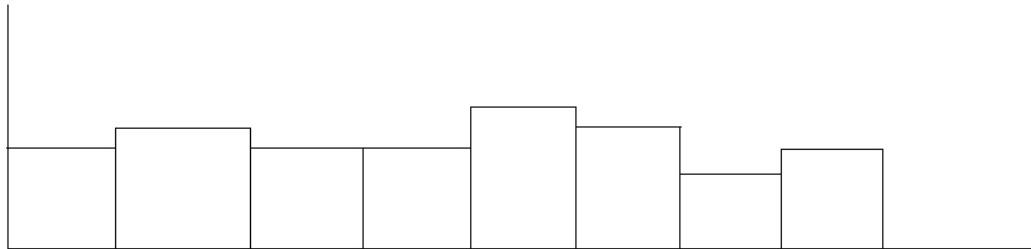


Figure 13.2

- A. uniform
- B. exponential
- C. normal
- D. chi-square

## Solutions to Exercises in Chapter 13

### Solution to Exercise 13.1 (p. 207)

- a.  $H_0: \mu_L = \mu_T = \mu_J$
- c.  $\text{df}(n) = 2; \text{df}(d) = 12$
- e. 0.67
- f. 0.5305
- h. Decision: Do not reject null; Conclusion: Means are same

### Solution to Exercise 13.3 (p. 208)

- c.  $\text{df}(n) = 4; \text{df}(d) = 4$
- e. 3.00
- f.  $2(0.1563) = 0.3126$
- h. Decision: Do not reject null; Conclusion: Variances are same

### Solution to Exercise 13.5 (p. 208)

- c.  $\text{df}(n) = 6; \text{df}(d) = 98$
- e. 1.69
- f. 0.1319
- h. Decision: Do not reject null; Conclusion: Average lap times are the same

### Solution to Exercise 13.7 (p. 209)

- a.  $H_0: \mu_d = \mu_n = \mu_h = \mu_c$
- b. At least one average is different
- c.  $\text{df}(n) = 3; \text{df}(d) = 16$
- e. 8.69
- f. 0.0012
- h. Decision: Reject null; Conclusion: At least one average is different

### Solution to Exercise 13.11 (p. 209)

REVIEW QUESTION 1 Solution : 0.02

### Solution to Exercise 13.12 (p. 209)

REVIEW QUESTION 2 Solution : 0.40

### Solution to Exercise 13.13 (p. 210)

REVIEW QUESTION 3 Solution :  $\frac{100}{140}$

### Solution to Exercise 13.14 (p. 210)

REVIEW QUESTION 4 Solution  $\frac{10}{60}$

### Solution to Exercise 13.15 (p. 210)

REVIEW QUESTION 5 Solution : p-value = 0 ; Reject null hypothesis ; Conclude dependent events

### Solution to Exercise 13.16 (p. 210)

REVIEW QUESTION 6 Solution : 8.4

### Solution to Exercise 13.17 (p. 210)

REVIEW QUESTION 7 Solution :  $B(14, 0.60)$

### Solution to Exercise 13.18 (p. 210)

REVIEW QUESTION 8 Solution : D

### Solution to Exercise 13.20 (p. 210)

REVIEW QUESTION 10 Solution : 0.6321

### Solution to Exercise 13.21 (p. 210)

REVIEW QUESTION 11 Solution : \$360

### Solution to Exercise 13.22 (p. 211)

REVIEW QUESTION 12 Solution :  $N\left(72, \frac{72}{\sqrt{5}}\right)$

**Solution to Exercise 13.23 (p. 211)**

REVIEW QUESTION 13 Solution : p-value = 0.0006; Reject null; Conclude averages are not equal

**Solution to Exercise 13.24 (p. 211)**

REVIEW QUESTION 14 Solution : p-value = 0; Reject null; Conclude proportion of males is higher

**Solution to Exercise 13.25 (p. 211)**

REVIEW QUESTION 15 Solution : D

**Solution to Exercise 13.26 (p. 211)****REVIEW QUESTION 16 Solution**

- a. No
- b. Yes,  $P(M \text{ and } 30+) = 0$

**Solution to Exercise 13.27 (p. 212)**

REVIEW QUESTION 17 Solution :  $\frac{12}{38}$

**Solution to Exercise 13.28 (p. 212)**

REVIEW QUESTION 18 Solution : No; p-value = 0

**Solution to Exercise 13.29 (p. 212)**

REVIEW QUESTION 19 Solution : A

# **Chapter 14**

# **Appendix**

## **14.1 Data Sets<sup>1</sup>**

### **14.1.1 Lap Times**

The following tables provide lap times from Terri Vogel's Log Book. Times are recorded in seconds for 2.5-mile laps completed in a series of races and practice runs.

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m17132/1.4/>>.

**Race Lap Times (in Seconds)**

	<b>Lap 1</b>	<b>Lap 2</b>	<b>Lap 3</b>	<b>Lap 4</b>	<b>Lap 5</b>	<b>Lap 6</b>	<b>Lap 7</b>
Race 1	135	130	131	132	130	131	133
Race 2	134	131	131	129	128	128	129
Race 3	129	128	127	127	130	127	129
Race 4	125	125	126	125	124	125	125
Race 5	133	132	132	132	131	130	132
Race 6	130	130	130	129	129	130	129
Race 7	132	131	133	131	134	134	131
Race 8	127	128	127	130	128	126	128
Race 9	132	130	127	128	126	127	124
Race 10	135	131	131	132	130	131	130
Race 11	132	131	132	131	130	129	129
Race 12	134	130	130	130	131	130	130
Race 13	128	127	128	128	128	129	128
Race 14	132	131	131	131	132	130	130
Race 15	136	129	129	129	129	129	129
Race 16	129	129	129	128	128	129	129
Race 17	134	131	132	131	132	132	132
Race 18	129	129	130	130	133	133	127
Race 19	130	129	129	129	129	129	128
Race 20	131	128	130	128	129	130	130

**Table 14.1**

### Practice Lap Times (in Seconds)

	Lap 1	Lap 2	Lap 3	Lap 4	Lap 5	Lap 6	Lap 7
Practice 1	142	143	180	137	134	134	172
Practice 2	140	135	134	133	128	128	131
Practice 3	130	133	130	128	135	133	133
Practice 4	141	136	137	136	136	136	145
Practice 5	140	138	136	137	135	134	134
Practice 6	142	142	139	138	129	129	127
Practice 7	139	137	135	135	137	134	135
Practice 8	143	136	134	133	134	133	132
Practice 9	135	134	133	133	132	132	133
Practice 10	131	130	128	129	127	128	127
Practice 11	143	139	139	138	138	137	138
Practice 12	132	133	131	129	128	127	126
Practice 13	149	144	144	139	138	138	137
Practice 14	133	132	137	133	134	130	131
Practice 15	138	136	133	133	132	131	131

**Table 14.2**

#### 14.1.2 Stock Prices

The following table lists initial public offering (IPO) stock prices for all 1999 stocks that at least doubled in value during the first day of trading.

### IPO Offer Prices

\$17.00	\$23.00	\$14.00	\$16.00	\$12.00	\$26.00
\$20.00	\$22.00	\$14.00	\$15.00	\$22.00	\$18.00
\$18.00	\$21.00	\$21.00	\$19.00	\$15.00	\$21.00
\$18.00	\$17.00	\$15.00	\$25.00	\$14.00	\$30.00
\$16.00	\$10.00	\$20.00	\$12.00	\$16.00	\$17.44
\$16.00	\$14.00	\$15.00	\$20.00	\$20.00	\$16.00
\$17.00	\$16.00	\$15.00	\$15.00	\$19.00	\$48.00
\$16.00	\$18.00	\$9.00	\$18.00	\$18.00	\$20.00
\$8.00	\$20.00	\$17.00	\$14.00	\$11.00	\$16.00
\$19.00	\$15.00	\$21.00	\$12.00	\$8.00	\$16.00
\$13.00	\$14.00	\$15.00	\$14.00	\$13.41	\$28.00
\$21.00	\$17.00	\$28.00	\$17.00	\$19.00	\$16.00
\$17.00	\$19.00	\$18.00	\$17.00	\$15.00	
\$14.00	\$21.00	\$12.00	\$18.00	\$24.00	
\$15.00	\$23.00	\$14.00	\$16.00	\$12.00	
\$24.00	\$20.00	\$14.00	\$14.00	\$15.00	
\$14.00	\$19.00	\$16.00	\$38.00	\$20.00	
\$24.00	\$16.00	\$8.00	\$18.00	\$17.00	
\$16.00	\$15.00	\$7.00	\$19.00	\$12.00	
\$8.00	\$23.00	\$12.00	\$18.00	\$20.00	
\$21.00	\$34.00	\$16.00	\$26.00	\$14.00	

**Table 14.3**

NOTE: Data compiled by Jay R. Ritter of Univ. of Florida using data from Securities Data Co. and Bloomberg.

## 14.2 English Phrases Written Mathematically<sup>2</sup>

### 14.2.1 English Phrases Written Mathematically

When the English says:	Interpret this as:
X is at least 4.	$X \geq 4$
XThe minimum is 4.	$X \geq 4$
X is no less than 4.	$X \geq 4$
X is greater than or equal to 4.	$X \geq 4$
X is at most 4.	$X \leq 4$
XThe maximum is 4.	$X \leq 4$
Xis no more than 4.	$X \leq 4$
X is less than or equal to 4.	$X \leq 4$
Xdoes not exceed 4.	$X \leq 4$
Xis greater than 4.	$X > 4$
XThere are more than 4.	$X > 4$
Xexceeds 4.	$X > 4$
Xis less than 4.	$X < 4$
XThere are fewer than 4.	$X < 4$
Xis 4.	$X = 4$
Xis equal to 4.	$X = 4$
Xis the same as 4.	$X = 4$
Xis not 4.	$X \neq 4$
Xis not equal to 4.	$X \neq 4$
Xis not the same as 4.	$X \neq 4$
Xis different than 4.	$X \neq 4$

Table 14.4

---

<sup>2</sup>This content is available online at <http://cnx.org/content/m16307/1.5/>.

### **14.3 Symbols and their Meanings<sup>3</sup>**

#### **Symbols and their Meanings**

---

<sup>3</sup>This content is available online at <<http://cnx.org/content/m16302/1.9/>>.

Chapter (1st used)	Symbol	Spoken	Meaning
Sampling and Data	$\sqrt{ }$	The square root of	same
Sampling and Data	$\pi$	Pi	3.14159... (a specific number)
Descriptive Statistics	$Q1$	Quartile one	the first quartile
Descriptive Statistics	$Q2$	Quartile two	the second quartile
Descriptive Statistics	$Q3$	Quartile three	the third quartile
Descriptive Statistics	$IQR$	inter-quartile range	$Q3-Q1=IQR$
Descriptive Statistics	$\bar{x}$	x-bar	sample mean
Descriptive Statistics	$\mu$	mu	population mean
Descriptive Statistics	$s s_x sx$	s	sample standard deviation
Descriptive Statistics	$s^2 s_x^2$	s-squared	sample variance
Descriptive Statistics	$\sigma \sigma_x \sigma x$	sigma	population standard deviation
Descriptive Statistics	$\sigma^2 \sigma_x^2$	sigma-squared	population variance
Descriptive Statistics	$\Sigma$	capital sigma	sum
Probability Topics	$\{ \}$	brackets	set notation
Probability Topics	$S$	S	sample space
Probability Topics	$A$	Event A	event A
Probability Topics	$P(A)$	probability of A	probability of A occurring
Probability Topics	$P(A   B)$	probability of A given B	prob. of A occurring given B has occurred
Probability Topics	$P(A or B)$	prob. of A or B	prob. of A or B or both occurring
Probability Topics	$P(A and B)$	prob. of A and B	prob. of both A and B occurring (same time)
<i>continued on next page</i>			

Probability Topics	$A'$	A-prime, complement of A	complement of A, not A
Probability Topics	$P(A')$	prob. of complement of A	same
Probability Topics	$G_1$	green on first pick	same
Probability Topics	$P(G_1)$	prob. of green on first pick	same
Discrete Random Variables	PDF	prob. distribution function	same
Discrete Random Variables	X	X	the random variable X
Discrete Random Variables	$X \sim$	the distribution of X	same
Discrete Random Variables	B	binomial distribution	same
Discrete Random Variables	G	geometric distribution	same
Discrete Random Variables	H	hypergeometric dist.	same
Discrete Random Variables	P	Poisson dist.	same
Discrete Random Variables	$\lambda$	Lambda	average of Poisson distribution
Discrete Random Variables	$\geq$	greater than or equal to	same
Discrete Random Variables	$\leq$	less than or equal to	same
Discrete Random Variables	=	equal to	same
Discrete Random Variables	$\neq$	not equal to	same
Continuous Random Variables	$f(x)$	f of x	function of x
<i>continued on next page</i>			

Continuous Random Variables	$pdf$	prob. density function	same
Continuous Random Variables	$U$	uniform distribution	same
Continuous Random Variables	$Exp$	exponential distribution	same
Continuous Random Variables	$k$	k	critical value
Continuous Random Variables	$f(x) =$	f of x equals	same
Continuous Random Variables	$m$	m	decay rate (for exp. dist.)
The Normal Distribution	$N$	normal distribution	same
The Normal Distribution	$z$	z-score	same
The Normal Distribution	$Z$	standard normal dist.	same
The Central Limit Theorem	CLT	Central Limit Theorem	same
The Central Limit Theorem	$\bar{X}$	X-bar	the random variable X-bar
The Central Limit Theorem	$\mu_x$	mean of X	the average of X
The Central Limit Theorem	$\mu_{\bar{x}}$	mean of X-bar	the average of X-bar
The Central Limit Theorem	$\sigma_x$	standard deviation of X	same
The Central Limit Theorem	$\sigma_{\bar{x}}$	standard deviation of X-bar	same
The Central Limit Theorem	$\Sigma X$	sum of X	same
The Central Limit Theorem	$\Sigma x$	sum of x	same
<i>continued on next page</i>			

Confidence Intervals	CL	confidence level	same
Confidence Intervals	CI	confidence interval	same
Confidence Intervals	EBM	error bound for a mean	same
Confidence Intervals	EBP	error bound for a proportion	same
Confidence Intervals	$t$	student-t distribution	same
Confidence Intervals	df	degrees of freedom	same
Confidence Intervals	$t_{\frac{\alpha}{2}}$	student-t with $\alpha/2$ area in right tail	same
Confidence Intervals	$\hat{p}$	p-prime; p-hat	sample proportion of success
Confidence Intervals	$\hat{q}$	q-prime; q-hat	sample proportion of failure
Hypothesis Testing	$H_0$	$H$ -naught, $H$ -sub 0	null hypothesis
Hypothesis Testing	$H_a$	$H$ -a, $H$ -sub a	alternate hypothesis
Hypothesis Testing	$H_1$	$H$ -1, $H$ -sub 1	alternate hypothesis
Hypothesis Testing	$\alpha$	alpha	probability of Type I error
Hypothesis Testing	$\beta$	beta	probability of Type II error
Hypothesis Testing	$\bar{X}_1 - \bar{X}_2$	$X_1$ -bar minus $X_2$ -bar	difference in sample means
	$\mu_1 - \mu_2$	$\mu$ -1 minus $\mu$ -2	difference in population means
	$P'_1 - P'_2$	$P$ 1-prime minus $P$ 2-prime	difference in sample proportions
	$p_1 - p_2$	$p$ 1 minus $p$ 2	difference in population proportions
Chi-Square Distribution	$\chi^2$	Ky-square	Chi-square
	$O$	Observed	Observed frequency
	$E$	Expected	Expected frequency
<i>continued on next page</i>			

Linear Regression and Correlation	$y = a + bx$	y equals a plus b-x	equation of a line
	$\hat{y}$	y-hat	estimated value of y
	$r$	correlation coefficient	same
	$\epsilon$	error	same
	$SSE$	Sum of Squared Errors	same
	$1.9s$	1.9 times s	cut-off value for outliers
F-Distribution and ANOVA	$F$	F-ratio	F ratio

**Table 14.5**

## Index of Keywords and Terms

**Keywords** are listed by the section with that keyword (page numbers are in parentheses). Keywords do not necessarily appear in the text of the page. They are merely associated with that section. *Ex.* apples, § 1.1 (1) **Terms** are referenced by the page they appear on. *Ex.* apples, 1

- A** accessibility, § (9)  
additional, § (9)  
adoption, § (9)  
appendix, § 14.1(215)
- B** Bernoulli, § 4.2(68)  
binomial, § 4.1(67), § 4.2(68)  
box, § 2.2(30)
- C** collaborative, § (1), § (5), § (9)  
collection, 5  
conditional, § 3.2(49)  
continuous, § 1.1(17), § 5.1(87), § 5.2(88),  
§ 5.3(93)  
cumulative, § 1.1(17)
- D** data, § 1.1(17), § 14.1(215)  
descriptive, § 2.2(30)  
deviation, § 2.2(30)  
discrete, § 1.1(17), § 4.1(67), § 4.2(68), § 4.3(78)  
distribution, § 4.3(78), § 5.1(87), § 5.2(88),  
§ 5.3(93)
- E** elementary, § (5), § (9), § (13), § 2.1(29),  
§ 3.1(49), § 3.2(49), § 3.3(60), § 4.1(67),  
§ 4.2(68), § 4.3(78), § 5.1(87), § 5.2(88),  
§ 5.3(93), § 6.1(99), § 6.2(99), § 6.3(105),  
§ 7.1(109), § 7.2(109), § 7.3(115), § 8.1(121),  
§ 8.2(121), § 8.3(131), § 9.1(139), § 9.2(140),  
§ 9.3(146), § 10.1(153), § 10.2(153), § 10.3(164),  
§ 11.1(169), § 11.2(170), § 11.3(177), § 12.1(187),  
§ 12.2(188), § 12.3(189), § 13.1(207), § 13.2(207),  
§ 13.3(209), § 14.1(215), § 14.2(219), § 14.3(220)  
elementary statistics, § (15), § 2.2(30)  
event, § 3.2(49)  
exclusive, § 3.2(49)  
exercise, § 2.2(30), § 4.2(68), § 4.3(78), § 5.3(93)  
exercises, § 3.3(60)  
exponential, § 5.1(87)
- F** formula, § 3.1(49), § 4.1(67), § 5.1(87)  
frequency, § 1.1(17), § 2.2(30)
- function, § 4.1(67), § 4.3(78), § 5.2(88), § 5.3(93)
- G** geometric, § 4.1(67), § 4.2(68)  
graph, § 5.2(88)  
guide, § (9)
- H** histogram, § 2.2(30)  
Homework, § 1.1(17), § 2.2(30), § 3.3(60),  
§ 4.2(68), § 4.3(78), § 5.2(88), § 5.3(93)  
hypergeometric, § 4.2(68)  
hypergeometrical, § 4.1(67)
- I** independent, § 3.2(49)
- M** mean, § 2.2(30)  
median, § 2.2(30)  
mode, § 2.2(30)  
modules, 5  
mutually, § 3.2(49)
- P** percentile, § 2.2(30)  
plot, § 2.2(30)  
Poisson, § 4.1(67), § 4.2(68)  
population, § 2.2(30), § 5.3(93)  
practice, § 3.3(60), § 4.3(78), § 5.3(93)  
probability, § 3.1(49), § 3.2(49), § 3.3(60),  
§ 4.1(67), § 4.3(78), § 5.1(87), § 5.3(93)  
problem, § 2.2(30), § 4.2(68)
- Q** qualitative, § 1.1(17)  
quantitative, § 1.1(17)  
quartile, § 2.2(30)
- R** random, § 1.1(17), § 4.1(67), § 4.2(68), § 4.3(78),  
§ 5.1(87), § 5.2(88), § 5.3(93)  
relative, § 1.1(17), § 2.2(30)  
resources, § (9)  
review, § 3.3(60), § 4.3(78), § 5.3(93)
- S** sample, § 1.1(17), § 2.2(30)  
sampling, § 1.1(17)  
set, § 14.1(215)  
standard, § 2.2(30)

- statistics, § (1), § (5), § (9), § (13), § 1.1(17),  
§ 2.1(29), § 2.2(30), § 3.1(49), § 3.2(49), § 3.3(60),  
§ 4.1(67), § 4.2(68), § 4.3(78), § 5.1(87), § 5.2(88),  
§ 5.3(93), § 6.1(99), § 6.2(99), § 6.3(105),  
§ 7.1(109), § 7.2(109), § 7.3(115), § 8.1(121),  
§ 8.2(121), § 8.3(131), § 9.1(139), § 9.2(140),  
§ 9.3(146), § 10.1(153), § 10.2(153), § 10.3(164),  
§ 11.1(169), § 11.2(170), § 11.3(177), § 12.1(187),  
§ 12.2(188), § 12.3(189), § 13.1(207), § 13.2(207),  
§ 13.3(209), § 14.1(215), § 14.2(219), § 14.3(220)  
summary, § 5.1(87)  
supplemental, § (9)
- survey, § 1.1(17)
- T** teacher, § (9)
- U** uniform, § 5.1(87)  
usage, § (9)
- V** variable, § 4.1(67), § 4.2(68), § 4.3(78), § 5.1(87),  
§ 5.3(93)  
variables, § 5.2(88)  
version, § (9)  
video, § (9)

## Attributions

Collection: Collaborative Statistics Homework Book: Custom Version modified by R. Bloom Edited by: Roberta Bloom URL: <http://cnx.org/content/col10619/1.2/> License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Collaborative Statistics Summary of Modifications by R. Bloom" Used here as: "Collaborative Statistics: custom version modified by R. Bloom" By: Roberta Bloom URL: <http://cnx.org/content/m18941/1.3/> Pages: 1-4 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Preface to "Collaborative Statistics"" Used here as: "Preface by S. Dean and B. Illowsky" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16026/1.16/> Pages: 5-7 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Collaborative Statistics: Additional Resources" Used here as: "Additional Resources" By: Barbara Illowsky, Ph.D., Susan Dean URL: <http://cnx.org/content/m18746/1.5/> Pages: 9-11 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Collaborative Statistics: Author Acknowledgements" Used here as: "Author Acknowledgements" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16308/1.9/> Page: 13 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Collaborative Statistics: Student Welcome Letter" Used here as: "Student Welcome Letter" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16305/1.5/> Pages: 15-16 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Sampling and Data: Homework (modified R. Bloom)" Used here as: "Homework (modified R. Bloom)" By: Roberta Bloom URL: <http://cnx.org/content/m18858/1.3/> Pages: 17-25 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: Sampling and Data: Homework By: Susan Dean, Barbara Illowsky, Ph.D., Roberta Bloom URL: <http://cnx.org/content/m18643/1.1>

Module: "Descriptive Statistics: Summary of Formulas" Used here as: "Summary of Formulas" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16310/1.8/> Pages: 29-30 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Descriptive Statistics: Homework (edited R. Bloom)" Used here as: "Homework (modified R. Bloom)" By: Roberta Bloom URL: <http://cnx.org/content/m18645/1.4/> Pages: 30-45 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: Descriptive Statistics: Homework By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16801/1.11>

Module: "Probability Topics: Summary of Formulas" Used here as: "Summary of Formulas" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16843/1.4/> Page: 49 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Probability Topics: Homework (modified R. Bloom)" Used here as: "Homework (modified R. Bloom)" By: Roberta Bloom URL: <http://cnx.org/content/m18924/1.2/> Pages: 49-60 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: Probability Topics: Homework By: Barbara Illowsky, Ph.D., Susan Dean URL: <http://cnx.org/content/m16836/1.8/>

Module: "Probability Topics: Review Questions" Used here as: "Review Questions" By: Roberta Bloom URL: <http://cnx.org/content/m19023/1.1/> Pages: 60-62 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: Probability Topics: Review By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16842/1.8/>

Module: "Discrete Random Variables: Summary of the Discrete Probability Functions" Used here as: "Summary of Functions" By: Susan Dean, Barbara Illowsky, Ph.D. URL:

<http://cnx.org/content/m16833/1.8/> Pages: 67-68 Copyright: Maxfield Foundation License:  
<http://creativecommons.org/licenses/by/2.0/>

Module: "Discrete Random Variables: Homework (modified R. Bloom)" Used here as: "Homework (modified R. Bloom)" By: Roberta Bloom URL: <http://cnx.org/content/m18927/1.2/> Pages: 68-78  
Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: Discrete Random Variables: Homework By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16823/1.9/>

Module: "Discrete Random Variables: Review Questions" Used here as: "Review Questions" By: Roberta Bloom URL: <http://cnx.org/content/m19021/1.1/> Pages: 78-81 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: Discrete Random Variables: Review By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16832/1.8/>

Module: "Continuous Random Variables: Summary of The Uniform and Exponential Probability Distributions" Used here as: "Summary of the Uniform and Exponential Probability Distributions" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16813/1.10/> Pages: 87-88 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Continuous Random Variables: Homework" Used here as: "Homework" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16807/1.12/> Pages: 88-93 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Continuous Random Variables: Review Questions" Used here as: "Review Questions" By: Roberta Bloom URL: <http://cnx.org/content/m19020/1.1/> Pages: 93-95 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: Continuous Random Variables: Review By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16810/1.9/>

Module: "Normal Distribution: Summary of Formulas" Used here as: "Summary of Formulas" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16987/1.4/> Page: 99 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Normal Distribution: Homework" Used here as: "Homework" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16978/1.20/> Pages: 99-105 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Normal Distribution: Review Questions" Used here as: "Review Questions" By: Roberta Bloom URL: <http://cnx.org/content/m19027/1.1/> Pages: 105-106 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: Normal Distribution: Review By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16985/1.7/>

Module: "Central Limit Theorem: Summary of Formulas" Used here as: "Summary of Formulas" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16956/1.6/> Page: 109 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Central Limit Theorem: Homework (modified R. Bloom)" Used here as: "Homework: CLT (modified R. Bloom)" By: Roberta Bloom URL: <http://cnx.org/content/m18940/1.2/> Pages: 109-115  
Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: Central Limit Theorem: Homework By: Barbara Illowsky, Ph.D., Susan Dean URL: <http://cnx.org/content/m16952/1.9/>

Module: "Central Limit Theorem: Review (modified R. Bloom)" Used here as: "Review Questions" By: Roberta Bloom URL: <http://cnx.org/content/m18863/1.2/> Pages: 115-117 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: Central Limit Theorem: Review By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16955/1.8/>

Module: "Confidence Intervals: Summary of Formulas" Used here as: "Summary of Formulas" By: Susan

Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16973/1.7> Page: 121 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Confidence Intervals: Homework" Used here as: "Homework" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16966/1.12> Pages: 121-131 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Confidence Intervals: Review Questions" Used here as: "Review Questions" By: Roberta Bloom URL: <http://cnx.org/content/m19018/1.1> Pages: 131-134 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: Confidence Intervals: Review By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16972/1.7>

Module: "Hypothesis Testing of Single Mean and Single Proportion: Summary of Formulas" Used here as: "Summary of Formulas" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16996/1.7> Page: 139 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Hypothesis Testing of Single Mean and Single Proportion: Homework (modified R. Bloom)" Used here as: "Homework (modified R. Bloom)" By: Roberta Bloom URL: <http://cnx.org/content/m18867/1.1> Pages: 140-146 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: Hypothesis Testing of Single Mean and Single Proportion: Homework By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17001/1.8>

Module: "Hypothesis Testing of Single Mean and Single Proportion: Review Questions" Used here as: "Review Questions" By: Roberta Bloom URL: <http://cnx.org/content/m19025/1.1> Pages: 146-148 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: Hypothesis Testing of Single Mean and Single Proportion: Review By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17013/1.7>

Module: "Hypothesis Testing: Two Population Means and Two Population Proportions: Summary of Types of Hypothesis Tests" Used here as: "Summary of Types of Hypothesis Tests" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17044/1.5> Page: 153 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Hypothesis Testing of Two Means and Two Proportions: Homework" Used here as: "Homework" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17023/1.15> Pages: 153-164 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>  
Module: "Hypothesis Testing of Two Means and Two Proportions: Review Questions" Used here as: "Review Questions" By: Roberta Bloom URL: <http://cnx.org/content/m19028/1.1> Pages: 164-165 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: Hypothesis Testing of Two Means and Two Proportions: Review By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17021/1.7>

Module: "The Chi-Square Distribution: Summary of Formulas" Used here as: "Summary of Formulas" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17058/1.5> Page: 169 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "The Chi-Square Distribution: Homework" Used here as: "Homework" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17028/1.18> Pages: 170-177 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "The Chi-Square Distribution: Review Questions" Used here as: "Review Questions" By: Roberta Bloom URL: <http://cnx.org/content/m19026/1.1> Pages: 177-181 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: The Chi-Square Distribution: Review By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17057/1.7>

Module: "Linear Regression and Correlation: 95% Critical Values of the Sample Correlation Coefficient

Table" Used here as: "95% Critical Values of the Sample Correlation Coefficient Table" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17098/1.5/> Pages: 187-188 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Linear Regression and Correlation: Summary" Used here as: "Summary" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17081/1.4/> Pages: 188-189 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Linear Regression and Correlation: Homework" Used here as: "Homework" By: Roberta Bloom URL: <http://cnx.org/content/m33266/1.1/> Pages: 189-203 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/3.0/> Based on: Linear Regression and Correlation: Homework By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17085/1.8/>

Module: "F Distribution and ANOVA: Summary" Used here as: "Summary" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17072/1.3/> Page: 207 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "F Distribution and ANOVA: Homework" Used here as: "Homework" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17063/1.9/> Pages: 207-209 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "F Distribution and ANOVA: Review Questions" Used here as: "Review Questions" By: Roberta Bloom URL: <http://cnx.org/content/m19022/1.1/> Pages: 209-212 Copyright: Roberta Bloom License: <http://creativecommons.org/licenses/by/2.0/> Based on: F Distribution and ANOVA: Review By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17070/1.7/>

Module: "Collaborative Statistics: Data Sets" Used here as: "Data Sets" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m17132/1.4/> Pages: 215-218 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

## ATTRIBUTIONS

Module: "Collaborative Statistics: English Phrases Written Mathematically" Used here as: "English Phrases Written Mathematically" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16307/1.5/> Page: 219 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

Module: "Collaborative Statistics: Symbols and their Meanings" Used here as: "Symbols and their Meanings" By: Susan Dean, Barbara Illowsky, Ph.D. URL: <http://cnx.org/content/m16302/1.9/> Pages: 220-225 Copyright: Maxfield Foundation License: <http://creativecommons.org/licenses/by/2.0/>

### **Collaborative Statistics Homework Book: Custom Version modified by R. Bloom**

This is a custom collection (by R. Bloom) of homework and review problems to accompany Collaborative Statistics textbook custom collection by R. Bloom. Content is derived from Collaborative Statistics written by Barbara Illowsky and Susan Dean, faculty members at De Anza College in Cupertino, California. The textbook by S. Dean and B. Illowsky was developed over several years and has been used in regular and honors-level classroom settings and in distance learning classes. This textbook is intended for introductory statistics courses being taken by students at two- and four-year colleges who are majoring in fields other than math or engineering. Intermediate algebra is the only prerequisite. The book focuses on applications of statistical knowledge rather than the theory behind it. This custom version of their collection has been modified by R. Bloom for her classes at De Anza College.

### **About Connexions**

Since 1999, Connexions has been pioneering a global system where anyone can create course materials and make them fully accessible and easily reusable free of charge. We are a Web-based authoring, teaching and learning environment open to anyone interested in education, including students, teachers, professors and lifelong learners. We connect ideas and facilitate educational communities.

Connexions's modular, interactive courses are in use worldwide by universities, community colleges, K-12 schools, distance learners, and lifelong learners. Connexions materials are in many languages, including English, Spanish, Chinese, Japanese, Italian, Vietnamese, French, Portuguese, and Thai. Connexions is part of an exciting new information distribution system that allows for **Print on Demand Books**. Connexions has partnered with innovative on-demand publisher QOOP to accelerate the delivery of printed course materials and textbooks into classrooms worldwide at lower prices than traditional academic publishers.